A potential nomenclature for the immunopolymorphism database (IPD) of chicken MHC genes: progress and problems

Hassnae Afrache[1], Clive A. Tregaskes[1] and Jim Kaufman[1,2,*]

[1]University of Cambridge, Department of Pathology, Tennis Court Road, Cambridge, CB2 1QP, U. K.

[2]University of Cambridge, Department of Veterinary Medicine, Madingley Road, Cambridge, CB2 0ES

*corresponding author, jfk31@cam.ac.uk

current email addresses

Hassnae Afrache, ha395@cam.ac.uk
Clive A. Tregaskes, ct383@cam.ac.uk
Jim Kaufman, jfk31@cam.ac.uk

Key words: BF-BL region, BF1, BF2, BLB1, BLB2, recombination

Abstract

Among the genes with the highest allelic polymorphism and sequence diversity are those encoding the classical class I and class II molecules of the major histocompatibility complex (MHC). Although many thousands of MHC sequences have been deposited in general sequence databases like GenBank, the availability of curated MHC sequences with agreed nomenclature has been enormously beneficial. Along with the immunogenetics-human leukocyte antigen (IMGT-HLA) database, a collection of databases for curated sequences of immune importance is available as the immunopolymorphism database (IPD). A recent addition is an IPD-MHC database for chickens. For many years, the nomenclature system for chicken MHC genes has been based on a list of standard, presumed to be stable, haplotypes. However, these standard haplotypes give different names to identical sequences. Moreover, the discovery of new recombinants between haplotypes and a rapid increase in newly-discovered alleles leaves the old system untenable. In this review, a new nomenclature is considered, for which alleles of different loci are given names based on the system used for other MHCs, and then haplotypes are named according to the alleles present. The new nomenclature system is trialed, first with standard haplotypes and then with validated sequences from the scientific literature. In the trial, some class II B sequences were found in both class II loci, presumably by gene conversion or inversion, so that identical sequences would receive different names. This situation prompts further suggestions to the new nomenclature system. In summary, there has been progress but also problems with the new IPD-MHC system for chickens.

Introduction

The creation of an HLA Database of curated class I and class II nucleotide sequences marked a pivotal moment for the human major histocompatibility complex (MHC) community, beginning a process that allowed all researchers to use common and agreed names for particular well-characterized sequences (Zemmour and Parham 1991; Marsh and Bodmer 1991; Robinson et al 2000). There have been many knock-on effects besides research into the MHC, such as providing the basis for single nucleotide polymorphisms (SNPs) to be used to impute MHC alleles for genome-wide association studies (GWAS), providing a model for other highly polymorphic immune gene systems such as the killer immunoglobulin-like receptors (KIRs), and providing the template for MHC systems in other species (Maccari et al 2018; Robinson et al 2015).

In addition to the database of human MHC sequences, now called the immunogenetics-human leukocyte antigen (IMGT-HLA), the immunopolymorphism database (IPD) hosts a variety of databases, each of which has expert curators to ensure that the sequences are validated and named according to appropriate nomenclature. Among these databases are those including MHC sequences (IPD-MHC) from non-human primates (NHP), cattle, swine, canines, ovines (sheep), caprines (goats), murids (rats) and chickens (Abraham et al 2018; Ballingall et al 2018). This review describes the progress made and the problems encountered in assembling, curating and extending the alleles of classical class I and class II B genes from the chicken MHC in order to properly implement the chicken database for IPD-MHC, efforts that continue to require serious re-consideration of the genetic nomenclature that has been in place for decades.

Discovery and first analyses of the chicken MHC

After the discovery of the mouse H-2 complex but before reports that led to the human HLA complex, a series of antigenic systems for chicken blood cells were described by Edwin Briles and co-workers (Briles et al 1950). These systems were discovered by injection of whole blood or blood cells from one chicken to another followed by haemagglutination to detect antibodies. As part of the effort to understand these serologically-detected systems, lines of chickens were bred to be homozygous for antigenic alleles of blood group B (Abplanalp et al 1981; Gilmour 1959), and functional assays emblematic of the MHC, including graft rejection, mixed lymphocyte reaction, graft-versus-host reaction, immune response to limited epitopes and autoimmunity, were found to be determined by the B locus (Bacon et al 1973; Gebriel and Nordskog 1983; Schierman and Nordskog 1961, 1963; Vilhelmova 1977).

Comparison of the patterns of serological reaction from different lines of chickens, along with absorbing populations of antibodies with various cell types from different lines, divided the B locus into a BG region that determined polymorphic erythrocyte antigens, and a BF-BL region that determined BL antigens found on lymphocytes and BF antigens on both erythrocytes and lymphocytes (Simonsen et al 1982). By immunoprecipitation and gel electrophoresis of radiolabelled molecules, the BF and BL antigens were found to the equivalent of class I and class II molecules respectively, while the BG antigens (also by this time called class IV antigens) were something else entirely (Wolf et al 1984). Interest in these regions was heightened by strong associations with economically-important diseases such as Marek's disease, for which particular B locus alleles, eventually located in the BF-BL region, conferred striking resistance or susceptibility (Briles et al 1977, 1983; Plachy et al 1992; reviewed in Miller and Taylor 2016).

The B locus was found with the ribosomal RNA genes of the nucleolar organizer region (NOR) on chicken chromosome 16 (Bloom and Bacon 1985), a microchromosome that is still in the process of being completely sequenced. After the seminal description of cosmid clones bearing class I and class II B genes (Guillemot et al 1988) followed by many studies utilizing molecular biology, our current understanding (Fig. 1a) has the B locus containing the classical MHC on the telomeric side of the long arm of this chromosome, followed by a region of repeats and the so-called Rfp-Y region containing non-classical class I and class II genes, then the NOR followed by regions containing scavenger receptor and olfactory receptor genes (reviewed in Kaufman 2013).

The B locus is now understood to include a multigene family of BG genes in the BG region, then a region containing many TRIM among other genes, then the BF-BL region containing the classical MHC genes, and at the end of the sequenced region a pair of CD1 genes, non-polymorphic non-classical class I genes that in mammals are located in an MHC-paralogous region on a different chromosome (Fig 1b). Within the BF-BL region are the polymorphic classical class I and class II B genes (Hosomichi et al 2008; Hunt and Fulton 1998; Jacob et al 2000; Kaufman et al 1999; Shaw et al 2007; Pharr et al 1998; Wallny et al 2006), along with the genes involved in antigen processing and peptide loading: DMA, DMB1, DMB2, tapasin, TAP1 and TAP2, all of which are also polymorphic (Atkinson et al 2000; Chazara et al 2011; Hosomichi et al 2008; van Hateren et al 2013; Walker et al 2005, 2011). In addition, several other genes are located in the BF-BL region: a BG gene (BG1), a natural killer (NK) receptor gene and a potential ligand (BNK and Blec), the transcription factor RING3 (Brd2, found in every MHC carefully examined), the complement component C4, a structural gene tenascin and the enzyme steroid hydroxylase. BG1 and BNK are known to be highly polymorphic (Chattaway et al 2016, Hosomichi et al 2008, Rogers et al 2008).

The first nomenclatures of the B locus

Nomenclature to describe B alleles began with the serological definition of standard haplotypes, starting with B1 (Briles et al 1982). As the complexity of the B locus was better appreciated and the molecular definition of genes and alleles progressed, this system was retained but with a variety of gene names followed by allele numbers. The currently published system of nomenclature (Miller et al 2004) recognizes 29 haplotypes and utilizes gene names based loosely on the original description from the cosmids of the B12 haplotype from the CB congenic chicken line along with later sequencing of the B locus cosmids (Guillemot et al 1988, Kaufman et al 1999), so the polymorphic class II B genes are called BLB1 and BLB2, and the polymorphic class I genes are called BF1 and BF2 (Fig. 1B). In addition, non-classical class I genes from the Rfp-Y region were found among the original cosmids and to be polymorphic (Afanassieff et al 2001; Miller et al 1994; Zoorob et al 1993), so these YF genes have also been incorporated into the published nomenclature system. The gene names are separated from the allele numbers by a star (or asterisk), with allele names coming from the surviving 20 standard B haplotypes but organized with allele groups in the first field separated by a colon from closely-related variants in the second field, based on the system for human MHC genes (as well as all other animals curated by the IPD-MHC, Abraham et al 2018; Ballingall et al 2018). Thus, the BF2 gene of the B2 haplotype (and for which only one sequence variant was described) has been named BF2*02:01. Many such sequence alleles are listed along with GenBank accession numbers and the chicken lines in which they were found (Miller et al 2004).

This currently accepted nomenclature system has been in place for 15 years, but there have been some difficulties in application. First, not all the sequences are known (Miller et al 2004). It would appear that B25 through B29 have not been analyzed at the molecular level. More seriously, the chicken lines for some standard haplotypes have apparently been lost without the sequences of their alleles being known (Miller et al 2004). So, the molecular identities of B1, B3, B10, B16, B20, B25, B27, B28 and B29 may never be known.

Second, the BLB and BF genes are apparently identical in some existing standard haplotypes. So, B4 and B13 differ in the BG region (the predominant determinant of the original serological identity) but are nearly identical in the sequenced genes of the BF-BL region (Hosomichi et al 2008; Hunt and Fulton 1998; Pharr et al 1998; Jacob et al 2000; Shaw et al 2007; Walker et al 2005; Wallny et al 2006).

Third, some published gene sequences for lines considered to have the same B haplotype are not the same, potentially due to issues of breeding, to nucleotide mis-incorporation during PCR or within the bacteria during cloning. For example, a sequence for a clone representing the dominantly-expressed class I gene from a B19 homozygote line of chickens was reported, followed later by a publication re-naming this sequence as B19var1, on the grounds that the "B19 type line" was held in a different institution and had a different sequence (Hunt et al 1994; Kaufman et al 1992). The second publication in fact used a congenic line derived from the type line, and subsequent analysis showed that the original sequence was correct (Hosomichi et al 2008), but not before the B19 and B19var1 names were used widely in the scientific literature (including Miller et al 2004). Similarly, some older sequences from clones derived from well-known lines do not agree with current sequences and may be due to inclusion of primer sequences during PCR or to nucleotide mis-incorporation during cloning (for example, Hunt and Fulton 1998; Liu et al 2002, Wallny et al 2006)

Fourth, recombination while rare does occur within the BF-BL region. The first example identified was the standard B19 haplotype, which by serology had BL in common with the B12 haplotype and BF closely-related to the B15 haplotype (Simonsen et al 1982). It is now clear by genomic sequences that the B19 haplotype is a hybrid of the B12 and B15 haplotypes, with the recombination site in the middle of the TAP gene and with a few nucleotide changes acquired in some genes (Hosomichi et al 2008; Jacob et al 2000; Rogers et al 2008; Shaw et al 2007; van Hateren et al 2013; Walker et al 2005, 2011; Wallny et al 2006). Several other examples among the standard haplotypes have since been identified, including recombination events apparently giving rise to the B5, B8 and B11 haplotypes found in various chicken flocks (Hosomichi et al 2008).

Fifth, comparison of genomic sequences for many BF-BL haplotypes identified long stretches of identical sequence between two haplotypes with completely different flanking sequences (Hosomichi et al 2008). These results were interpreted as "gene conversion", although in the absence of evidence about the homologous chromosome during meiosis, these could easily be due to "double reciprocal recombination" events rather than gene conversion as originally defined (Suyama et al 1959; Baltimore 1981). In any case, it has become clear that more complex events than just simple single recombination take place in the BF-BL region.

Finally, many sequences have been deposited in general sequence databases (like GenBank) that are not the same as the sequences known from the 29 standard haplotypes. Virtually all of these sequences come from PCR and only some have been controlled (typically by comparing results of independent PCRs) for nucleotide mis-incorporation or chimerism; only a few have been reported in publications. Moreover, there has been no clear mechanism to decide additional haplotype or allele numbers.

Of all these difficulties, only the nomenclature of recombinant haplotypes has been approached in the current system (Miller et al 2004). It was decided that the allele number of the BF2 sequence (rather than the serology primarily of the BG region) would be used to name the B locus haplotype, with recombinants given the designation "r" followed by a number in order of discovery. Thus, B2r1 would be the first recombinant described between the BF2 gene of the B2 haplotype but without information about the rest of the haplotype.

Given that recombination does occur within the BF-BL region, a more flexible and descriptive nomenclature might afford advantages. As a model, there is a long history of such nomenclature for the human MHC, in which recombination is relatively common and extended haplotypes are matter of great interest. The next sections describe the first attempts at such a nomenclature for the chicken MHC, as part of the effort to create an MHC-IPD database of agreed chicken MHC sequences with agreed names.

A potential new nomenclature for chicken MHC alleles and haplotypes

The question of a new nomenclature arose when considering how to curate allelic sequences for the chicken section of the IPD-MHC. It seemed likely that there would be many more sequences and haplotypes than the standard haplotypes examined up to now, and there was ample evidence for simple recombination and more complex events in the BF-BL region. Therefore, the current system of haplotypes followed by recombination numbers no longer seemed tenable, and a more descriptive and flexible system based on gene sequences was considered desirable.

The current method used for naming chicken MHC genes is not too distant from the nomenclature for human MHC genes, which is also proposed for MHC genes of other vertebrates (Maccari et al 2018; Miller et al 2004). As mentioned above for both systems, the gene name is followed by a star (asterisk), then the number of the allele group (also referred to as a "designation") is followed by a colon (often referred to as the "first field"), and then the number of a non-synonymous variant within the allele group is followed by a colon (the second field). In human nomenclature, a third field is the number of the synonymous variant within the variant group followed by a colon, and a fourth field gives the number of the variant in non-coding regions.

However, radical changes in the chicken nomenclature are envisaged for the naming of the allele groups and the definition of haplotypes. Instead of all genes within a defined haplotype having the number of that haplotype without regard to sequence, each gene (genetic locus) would have a list of sequence alleles and a haplotype would be named after the sequence alleles present in that haplotype. Thus, a particular sequence would have a single allele name regardless of haplotype. Also, closely-related sequences would have the same first field but differ in the second and/or third fields (and as more complete data is obtained, in the fourth fields), rather than reflect a haplotype number.

In deference to the decades of extensive and outstanding research into the chicken MHC, it seemed appropriate to treat the sequences in three groups: the sequences of the standard haplotypes for which the most information is available, the sequences outside of the standard haplotypes that are published with suitable controls and replication, and the sequences that are only found in existing general sequence databases or are currently being identified.

For the existing standard haplotypes, the process is to compare the sequences of each genetic locus between all haplotypes, then start with the lowest numbered haplotype to name identical and closely-related sequences appropriately with allele and variant names, then consider the haplotype with the next lowest number and so forth, naming any unique sequences with the allele number of that haplotype. Thus, in the absence of an available B1 haplotype, the sequences found in the B2 haplotype would be given the names BLB1*002:01:01, BLB2*002:01:01, BF1*002:01:01 and BF2*002:01:01.

For the sequences in the literature that are outside the standard haplotypes, almost all have entries in general sequence databases, some are described to be in haplotypes, and almost all are partial genes amplified by PCR from genomic DNA. Most usually, exon 2 from BLB genes and exons 2, 3 or both from BF genes are reported, but occasionally further exon or intron sequence is available; only those BF sequences that included both exon 2 and exon 3 would be considered. The first step is to ensure that independent amplifications have given the same results; if not, the sequence cannot be considered valid. As a second step with validated sequences, the gene from which the sequence was likely to have been derived is inferred, then this sequence is compared to the alleles of that genetic locus from the standard haplotypes, identical or closely-related sequences are named accordingly, and finally any unique sequences are given allele numbers starting with 30, well above the standard haplotypes for which sequences are available.

For those sequences that are only found in standard sequence databases such as GenBank (as well as those that are being discovered in an MHC typing effort described below), the first requirement is replication, validating only sequences found from independent amplifications (from independent experiments with the same bird, or from different birds, lines or studies). Sadly, most sequences in GenBank are present only once, and often the sequences in different entries from the same study differ only in one or two nucleotides and are most likely to be the result of nucleotide mis-incorporation during PCR for which no controls have been performed (Online Resource 1). Those sequences with replication are processed in the same way as the sequences from the literature.

Application of the potential new nomenclature to the standard B haplotypes

To carry out the process described above for the standard B haplotypes, two analyses based on alignments of both nucleotide and amino acid sequences were employed: phylogenetic trees to establish apparent clades and distance matrices to establish the number of sequence differences. The alignments were first carried out for exon 2 to exon 3 of the classical class I sequences (the peptide-binding domain without the intervening intron, which is anyway nearly invariant, Shaw et al 2007), and for exon 2 of the classical class II B sequences (without the rest of the gene, which is also nearly invariant, Jacob et al 2000), and later extended to the whole coding sequence (CDS equivalent) when possible. These alignments are relatively straightforward since there are no indels in the exons encoding the peptide-binding domains between any of the class I sequences or between any of the class II B sequences.

However, one might argue that a more sensible approach would be to compare alleles from a particular locus, rather than all class I sequences and all class II B sequences, given that differences in function might lead to differences in the sequence positions of variation. The biological functions of the polymorphic classical MHC loci, based on expression level and tissue distribution primarily from cDNA along with functional data mostly from *in vitro* cellular immunology, suggest that there is good reason to consider the alleles of these loci separately. Both class I loci are widely expressed, but the BF2 gene is strongly expressed and recognised by cytotoxic T lymphocytes (CTLs), while the BF1 gene is relatively poorly-expressed and recognised by NK cells (Ewald and Livant 2004; Fulton et al 1995; Kim et al 2018; Juul-Madsen et al 2000; Shaw et al 2007; Thacker et al 1995; Wallny et al 2006). In contrast, the class II genes differ in their tissue distribution: the BLB2 gene is strongly expressed in many tissues, but the BLB1 gene is poorly-expressed except in the intestine (Jacob et al 2000; Parker et al 2017; Pharr et al 1998). The assignment of particular sequences to the polymorphic classical MHC loci was possible for many of the standard haplotypes, based on complete BF-BL region sequences (Hosomichi et al 2008; Kaufman et al 1999) along with supporting data from complete or partial cDNA and gene sequences available from one or more chicken lines (Online Resource 2), so lists of alleles were established for each genetic locus separately.

To establish apparent clades that could be considered allele groups of variants for the first and second fields of the name, phylogenetic trees were used. Both nucleotide and amino acid sequences were used, of which examples for the amino acids from the α1 and α2 domains (and nucleotides for exons 2-3) of BF and from the β1 domains (exon 2) of BLB sequences are shown (Fig. 2; Online Resource 3).

Typically all the BF sequences from a particular B haplotype are identical, but not necessarily all BF sequences for a particular clade (Fig. 2a; Online Resource 3a). For instance, the five BF2 amino acid sequences from the B2 haplotype in the top clade are identical despite originating from multiple chicken lines and being analysed by different researchers in different ways; the same is true for the nucleotide sequences. Similarly, all the BF2 sequences from each of the B6, B12, B17 and B21 haplotypes and the BF1 sequences from each of the B2 and B6 haplotypes are identical within their clades, both as amino acids and as nucleotides. The two BF2 sequences from the B14 haplotype differ in one nucleotide leading to one amino acid difference; this appears to be due to a cloning artefact. However, other clades have sequences from more than one B haplotype. For instance, the BF2 sequences from the B4 and B13 haplotypes are identical in exons 2 and 3 at the amino acid and nucleotide levels. Similarly, BF2 sequences from the B15 haplotype are identical as are those from the B19 haplotype, but together they form a clade of closely-related sequences. The same is true for both the BF2 sequences of the B5, B8 and B11 haplotypes. Very striking is the BF1 clade of B4, B13, B15, B21 and B24 sequences, which are identical in these exons, although they can differ slightly in other exons. Only single sequences are available for BF1 from the B9, B17 and B23 haplotypes, as well as BF2 from the B9, B18, B23 and B24 haplotypes. In fact, the BF1 sequence from the B17 haplotype appears on its own just outside the B4 clade, but was considered to be part of the B4 allelic group based on distances matrices, as discussed below. Similarly, the BF2 sequences from the B2 and B23 haplotypes seem to be in different clades for amino acids, but arguably in the same clade for nucleotides; the decision to consider them as the same allelic group was based on distance matrices.

Overall, there are fewer clades for BF1 than BF2 sequences, and there are no BF1 sequences clustered in the same narrow clade with BF2 sequences in this tree. However, in some analyses not shown here, the BF sequences form two much larger clades: most BF1 sequences in one large clade and all the BF2 sequences along with the BF1 sequences from the B2 and B9 haplotypes in the other large clade. Thus, on the basis of sequence alone, there would be a reasonable probability of assigning a newly-discovered BF sequence that is not closely-related to an existing sequence to the right genetic locus.

The BLB sequences overall show similar features, except that BLB1 and BLB2 sequences are more intermixed (Fig. 2b; Online Resource 3b). For haplotypes with multiple sequences in the literature, they are identical for each haplotype, but only BLB1 from B15 and BLB2 from B2, B15 and B21 each form their own clades. Some clades are only represented by a single sequence: BLB1 from B17, and BLB2 from B9, B14, B17 and B23. All other sequences are in clades with several haplotypes. Clusters of BLB1 sequences are found for B2, B6 and B8, for B4, B13, B21 and B24, for B5 and B23, for B9, B11 and B14, and for B12 and B19. Clusters of BLB2 sequences are found for B4 and B13, for B5, B6 and B11, for B8 and B24, and for B12 and B19. There are no instances of BLB1 and BLB2 sequences being closely-related in a well-supported clade, but several examples of more distantly-related BLB1 and BLB2 sequences together in moderately-supported clades. Thus, on the basis of sequence alone, it may become problematic to assign a newly-discovered BLB sequence to the right genetic locus, unless it is very closely-related to an existing sequence. In fact, this possibility turned out to be even more of a problem than anticipated, as described in more detail below.

Once the allele groups were established based on clades, they were analysed by distance matrices, starting with amino acids (Fig. 3) and then nucleotides (Online Resource 4). Initially, only exon 2 (β1 domain) of BLB sequences, and exons 2 and 3 (α1 and α2 domains) of BF sequences were compared, given that these regions determine the peptide-binding specificities. However, there are at least three concerns with this approach, which were rationalised as follows. First, not all the polymorphic positions within these exons/domains contribute to peptide-binding, but it was judged that the number of variable positions outside of the peptide-binding groove was likely to be small and therefore not affect assignments too much. Second, some variation in the whole gene/molecule can be found outside of these peptide-binding exons/domains, but it was judged that many of the available sequences outside of the standard haplotypes would not have more sequence than these exons/domains, and after the initial characterisation further adjustments based on the few additional differences could be made. Third, the relationships of recombinant alleles (for instance, a BF gene composed of exon 2 from one BF allele and exon 3 from another BF allele) would not be captured, but it was judged that there was no simple way to present this information in a name, and that further characterisation might compare BF exon 2/α1 and exon 3/α2 sequences separately to look for interesting differences. In any case, after the initial examination, the analyses were extended to the whole protein (Online Resource 5).

Overall it turned out to be relatively easy to make sensible assignments for the standard haplotypes, based on these difference matrices. Most of the BF sequences differed by over 20 amino acids, while most of the BLB sequences differed by at least 12 amino acids. A few sequences for each were identical (green highlights in Fig. 3), and a few more had only one or two amino acid differences (blue highlights). The difficult judgement was how many more differences were too many to be a variant within an allele group. As for human alleles, four amino acid differences per domain (totalled to eight for BF sequences) were used as a cut-off (yellow highlights). This felt to be reasonable, since the BF2 molecule from the B19 haplotype is known to be derived from the B15 haplotype, and differs in seven amino acids over the α1 and α2 domains. Those comparisons with more than five but less than ten differences per exon were considered not to be variants of each other, but to warrant further analysis in the future.

Once the alleles and variants were established, then they were named in numerical order of haplotype. Illustrations of those decisions for the standard haplotypes are given for the whole amino acid coding sequence (CDS) (Fig. 4a) and for the peptide-binding domains (Fig. 4b), in which unique sequences have no colour, identical sequences are coloured, and close variants have the same colour but are striped. Given that there are no sequences available for the B1 haplotype (Miller et al 2004) and the type line is no longer in existence (S. Lamont, personal communication), the first standard haplotype known to be available is B2, for which there are sequences from multiple lines and sources. Thus, all the genes would receive the same allele number reflecting that basal haplotype, for example BLB1*002:01:01. There are again no B3 sequences or chicken lines available, but the sequences for B4 are not similar to B2, and the sequences for B5 are not similar to either B2 or B4. Thus, the allele numbers would reflect these haplotypes, for example BLB1*004:01:01 and BLB1*005:01:01.

The situation becomes more complicated for haplotypes with higher B numbers (Fig. 4a, b). For the B6 haplotype, the BF1 and BF2 genes are different from B2, B4 and B5, but the BLB1 sequence is identical to B2 throughout the coding sequence and is thus named BLB1*002:01:01, while the BLB2 sequence is identical to B5 throughout the coding sequence and is thus named BLB2*005:01:01. For the B8 haplotype, BLB1 is identical to B2 (and also B6) and so it would be named BLB1*002:01:01, BLB2 is not like any of the haplotypes with a lower number (but is identical to BLB2 from B24) and would be named BLB2*008:01:01, and both BF1 and BF2 are amino acid variants of the B5 genes (and are identical to the B11 genes) and would be named BF1*005:02:01 and BF2*005:02:01. The rest of the haplotypes for which there are sequences available would be named in a similar way, except that the B14 and one B15 haplotype have null alleles for BF1, while another B15 haplotype has BF1*004:02:01. All four genes of the B13 haplotype are identical to B4 in the peptide-binding region (although the BLB2 sequence for B13 differs in one nucleotide leading to one amino acid change in exon 3), so they all would have names that reflect this fact, giving the haplotype BLB1*004:01:01-BLB2*004:02:01-BF1*004:01:01-BF2*004:01:01. Also, the BF2 gene from the B19 haplotype is extremely similar to the B15 haplotype, and the other three genes are identical or very similar to the B12 haplotype, so this haplotype becomes BLB1*012:01:01-BLB2*012:02:01-BF1*012:02:01-BF2*015:02:01.

Overall, it can be seen that most haplotypes are patchworks of identical or similar sequences shared with other haplotypes, suggesting recombination or other processes over a considerable period of time. Moreover, some allelic groups have many variants, particularly obvious for the BF1*004 clade, although there are additional differences outside of exons 2 and 3 for some variants (comparing Fig. 4a with Fig. 4b). This is not a drawback for the standard haplotypes, for which there are complete gene sequences for all alleles, but becomes a difficulty below when comparing sequences from the literature, for which typically only partial sequences are available.

A serious drawback to this new nomenclature is that it is quite cumbersome. A convenient shorthand for the complicated gene name might be to leave out the last fields that are identical to the first variant described (so that BF1*002:01:01 would be simply "2", while BF1*005:02:01 would be "5:02"), and then present the haplotype as a string in the order BLB1-BLB2-BF1-BF2. Thus, the B2 haplotype would be 2-2-2-2, the B6 haplotype 2-5-6-6, the B8 haplotype 2-8-5:02-5:02, and the B19 haplotype 12-12:02-12:02-15:02. An even greater simplification might be to name the BF-BL haplotypes after the BF2 genes (as has been done in Miller et al 2004), but as "Bfbl haplotypes" to distinguish them from B haplotypes that describe the whole B locus. Thus, the haplotypes above would be named Bfbl 2, 6, 5:02 and 15:02.

Application of the potential new nomenclature to other validated sequences from the literature

In order to extend the lists of alleles and haplotypes beyond the standard haplotypes, an extensive search of the general sequence database and of the scientific literature was carried out (Online Resources 1, 2). Most of these sequences came from PCR amplifications of BLB exon 2 and of BF exon 2 to exon 3 from genomic DNA, subsequently cloned and sequenced. For some sequences, such amplifications from cDNA were also available. A few sequences were longer, but there are only two publications with sequences of complete genes, one from a line with a B6 haplotype (Suzuki et al 2012) and the other from a line with a red junglefowl haplotype considered nearly identical with the B21 haplotype (Shiina et al 2007). In general, evidence for multiple independent isolations (independent PCRs from a single chicken, more than one individual chicken or line, or from different laboratories) was required for the sequence to be considered valid.

Only a few publications met the criteria of repeatability. Chief among them is a series of papers from the lab of Sandra Ewald (Li et al 1997, 1999; Livant et al 2001, 2004; supplemented with some direct submissions to GenBank for BLB1 sequences) that examined amplification of class I and class II B sequences from both cDNA and genomic DNA derived from lines of commercial broiler (that is, meat-type) chickens. A publication amplifying class I sequences from blue egg Caipira chickens in Brazil (Lima-Rosa et al 2004) reported many of the same sequences. In addition, one publication reported class II B sequences from three Chinese lines (Chen et al 2012) and another publication reported both class I and class II B sequences from a population of captive red junglefowl (Worley et al 2008). Other publications and GenBank entries lacked replication, and for some sequences there was clear confusion between the publication and the associated database entries (for example, some sequences from Chen et al 2012 and Worley et al 2008). Once the sequences were considered valid, they were compared to those in the standard haplotypes (Fig. 5; Online Resource 2). Some of these sequences are identical or very similar to sequences in the standard haplotypes (leading to differences in the colours between Fig. 4b and Fig. 5a).

Some of these class I and class II B sequences were (or could be) assembled into BF-BL haplotypes, which will be referred to below by provisional names (Fig. 5b; Online Resource 2). A few of these haplotypes are exactly as known for sequences from standard haplotypes, such as the Bfbl 2 (2-2-2-2) and 21 haplotypes (4-21-4:02-21). The gene sequences from the WLA line were reported as a B6 haplotype (Suzuki et al 2012), but it is related to the standard haplotype from the line GB-2 haplotype (Hosomichi et al 2008) by apparent recombination (2-8-6-6 compared to the standard haplotype 2-5-6-6), prompting a provisional name of Bflb 6b. It has not escaped the authors that such a name is not enormously different from the current method of naming recombinants with “r” and a number. Other full haplotypes assembled from sequences in the literature had BF2 alleles that are not found in the standard haplotypes (Fig. 5b). These include Bfbl 9:02 (4:03-33-4:04-9:02) and 30 (30-30-6:02-30) found in commercial and Brazilian chickens, 31 (31-31-31-31) found only in commercial chickens, 32 (32-32-4:02-32) found in commercial and wild chickens, 33 (9-34-null-33) in commercial, Brazilian and wild chickens, 38 [(109-109)-23:02-38], 39 (5-5:02-4-39) and 40 (33-35-23:03-40) found only in wild chickens.

For some studies in the literature, only partial haplotypes could be assembled or only single genes were reported (Fig. 5b). Partial haplotypes include Bfbl 9 (?-?-9:02-9) from Brazilian chickens, 17:02 (?-17-30:01:02-17:02) from commercial and Brazilian chickens, 17:03 (?-?-17-17:03) from Brazilian chickens, 24b (?-36-30-24) from commercial chickens, 34 [4-?-23-34] in Brazilian and wild chickens, 36 (?-?-23-36) from commercial chickens, and 36:02 (?-?-6-36:02) from Brazilian chickens. Some singletons are identical to sequences from standard haplotypes, such as BF1*004:01:01, BF2*004:01:01, BF2*014:01:01 and BF2*015:01:01. Other singletons are closely-related to standard sequences, such as BF2*014:02:01 and BF2*015:03:01 from commercial chickens. Still others are not related to known sequences, such as BF2*035:01:01 found in commercial and Brazilian chickens and BF2*037:01:01 found in Brazilian chickens.

As the closely-related sequences are reported to have been replicated, it is likely that they are real rather than some nucleotide mis-incorporation. All of the new Bfbl haplotypes except for 36:02, 37, 39 and 40 have been extended and/or amply verified in a wide-ranging typing exercise of commercial egg-layers and broilers, fancy breeds and local (indigenous) chickens (C. Tregaskes, R. Martin, H. Afrache and J. Kaufman, unpublished).

Some of these new haplotypes highlighted unexpected difficulties, which have become ever more prominent in the wide-ranging typing exercise mentioned above. The Bfbl 33 haplotype lacks a BF1 allele (apparently, since absence of evidence is not evidence of absence), but this is no longer unexpected since standard B14 and B15 haplotypes also can lack a BF1 allele at both genomic and cDNA levels (Wallny et al 2006; Shaw et al 2007). However, the two BLB sequences are found in clades with both BLB1 and BLB2 sequences; between gene PCRs established the locus for each of these alleles (Worley et al 2008). More confusingly, the Bfbl 34 haplotype from red junglefowl has only one BLB sequence that is identical to BLB1*004:01:01, and between-gene PCRs located this sequence in the BLB1 locus (Worley et al 2008). On the other hand, a Bfbl 34 haplotype has been found in the wide-ranging typing exercise that has two sequences neither of which is closely-related to particular BLB sequences (C. Tregaskes, R. Martin, H. Afrache and J. Kaufman, unpublished), one located in the BLB1 locus and the other in the BLB2 locus (F. Filaire, H. Afrache, C. Tregaskes and J. Kaufman, unpublished). Similarly, the Bfbl 38 haplotype has only one BLB sequence which again is not closely-related to any particular known BLB sequences, and between-gene PCRs show that this sequence to be present in the BLB1 locus between Blec and tapasin as well as in the BLB2 locus between Brd2 and tapasin (Worley et al 2008). Follow-up studies confirm that the same exon 2 sequence can be present in both the BLB1 and BLB2 loci (F. Filaire, R. Martin, H. Afrache, C. Tregaskes and J. Kaufman, unpublished), and a temporary assignment of such BLB sequences without a clear genetic location to numbers starting with 101 was established, along with the temporary use of curved parentheses to show that the location is unclear [such as "(109-109)" mentioned above]. As discussed in the section about phylogenetic trees, BLB1 and BLB2 clades are more intermixed than BF1 and BF2 clades, which may reflect this phenomenon.

It seems likely that the basis of the presence of (the same) BLB1 sequences in both the BLB1 and BLB2 genetic loci is due the compact nature of the BF-BL region and the fact that the BLB1 and BLB2 genes are in opposite transcriptional orientation. Gene conversion between homologous genes is thought to increase in frequency with decreasing physical distance between them (McCormack and Thompson 1990; Sayegh et al 1999). Also, recombination between genes in opposite transcriptional orientation leads to inversion (Lundqvist et al 2001; Zhao et al 2000) rather than deletion as found for genes in the same orientation (Fig. 6).

Discussion

An IPD-MHC database requires sequences that are validated and curated, but also requires a nomenclature that precisely allows the sequence to be identified in a convenient and biologically-meaningful way. The system used for human MHC sequences (based on genetic loci, allele groups and variants within those groups) was adapted for the chicken MHC sequences long ago (Miller et al 2004), but the allele names were based on traditional "standard" MHC haplotypes rather than on sequences. While recombination between haplotypes had been understood for many years (Simonsen et al 1982), recombination among haplotypes is now known to have been too prevalent (Hosomichi et al 2008) for a haplotype-based nomenclature to be sustained.

Some properties of the chicken MHC should make development of a nomenclature particularly easy. The BF-BL region is simple, with only two classical class I and two classical class II B genes that are polymorphic (Hosomichi et al 2008; Jacob et al 2000; Kaufman et al 1999; Shaw et al 2007; Wallny et al 2006). Unlike some species, there is very little copy number variation (CNV): BF1 null alleles and a third BLB gene B12c found in some B12 haplotypes (Shaw et al 2007; Zoorob et al 1993). The β_2-microglobulin and class II A (BLA) genes that encode the partner chains are non-polymorphic (Riegert et al 1996; Salomonsen et al 2003). The non-classical class I and class II B genes from the Rfp-Y region are sufficiently distant in sequence not to be amplified by typical PCR primers (Afanassieff et al 2001; Zoorob et al 1993). There is very little variation outside of the exons encoding the peptide-binding domains of the classical genes, including nearly-invariant introns in between (Jacob et al 2000; Shaw et al 2007). Also, the compact nature of the BF-BL region, in which most introns are very small compared to most jawed vertebrates, is well-suited for PCR- and sequence-based typing methods (Potts 2016; C. Tregaskes, R. Martin, H. Afrache and J. Kaufman, unpublished).

However, the compact and simple nature of the chicken MHC also has disadvantages for analysis. In particular, the rarity of recombination leads to relatively stable haplotypes that encouraged a particularly simple nomenclature, but there is enough recombination (Hosomichi et al 2008) to make this simple method untenable in the long run. On the flip side, the lack of recombination leads to difficulty in determining which gene within a haplotype is responsible for a biological trait. A newly-appreciated difficulty is the exchange of information between the class II B genes (Worley et al 2008), so that it may be difficult to assign a new sequence to a particular genetic locus. In fact, this kind of "concerted evolution" was first noticed long ago in the class II B genes of a closely-related species, the pheasant (Wittzell et al 1999). This exchange may be by gene conversion (which is increased by close proximity) or by inversion (which is possible due to the genes being in opposite transcriptional orientation) (McCormack and Thompson 1990; Lundqvist et al 2001; Sayegh et al 1999; Zhao et al 2000). In comparison, such exchange of sequence between the class I genes seems rare, with only the BF1 genes of the B2 and B9 haplotypes looking very similar to BF2 alleles, including peptide-binding motif for BF1*002:01 (Chappell et al 2015). However, the presence of "B4 minor" class I sequences amplified preferentially from cDNA of the BA5 and BA12 haplotypes (Li et al 1999) may mean that BF1*004 sequences can be highly expressed from the BF2 locus.

Having genes in opposite transcriptional orientation means that homologous recombination between them leads to inversion rather than loss of gene by deletion or CNV by unequal crossing-over (such as is seen for BG genes in the BG region, Salomonsen et al 2014). Indeed, the BF-BL region has several such gene pairs in opposite transcriptional orientation (BLB1/BLB2, TAP1/TAP2, BF1/BF2, BNK/Blec), which may have evolved to avoid loss of genes from what has been characterised as a "minimal essential MHC" (Kaufman et al 1995, 1999), which cannot afford to lose any "essential" genes.

The fact that the regulatory regions of these genes (for instance, promoters and 3' untranslated regions) are generally separate from the coding regions and can be independently exchanged may have led to changes in expression that up to now have been considered to be a property of a particular locus (for instance, high and wide expression from the BLB1 and BF1 loci rather than from the BLB2 and BF2 loci). Although the compact nature of chicken MHC genes means that relatively simple sequencing can be performed for the exons encoding the peptide-binding regions (at least compared to many other animal species with long introns), it is not yet routine to sequence long stretches of DNA from many individual chickens, so locating the sequences to particular genetic loci and determining the regulation of those sequences remains a challenge. However, such determination would seem to be essential to ensure clear assignments in the database.

This exchange of information between genes also leads to difficulties in the proposed new nomenclature system, which has as a central tenent that alleles are identified by sequence rather than by haplotype, so a unique sequence would have a unique name. For instance, the JF9 sequence was found in both the BLB1 and BLB2 loci of red junglefowl (Worley et al 2008; F. Filaire, R. Martin, H. Afrache, C. Tregaskes and J. Kaufman, unpublished), so should these identical exon 2 sequences get different names? The typing exercise mentioned above (C. Tregaskes, R. Martin, H. Afrache and J. Kaufman, unpublished) has found many examples of similar sequences in both loci, so it is a real difficulty. One potential modification to ameliorate this confusion would be to give each sequence a unique name, perhaps ensuring that (as much as possible) those sequences predominantly found in BLB1 loci have odd numbers for alleles, and those found in BLB2 loci have even numbers for alleles. The extent to which this is feasible has yet to be ascertained. If such a modification is implemented for the class II B loci, then perhaps (in the interest of consistency) it should be considered for the class I loci, despite the lower level of exchange between the BF1 and BF2 genes.

Another unexpected concern involves the assignment of variants to alleles groups. For the standard haplotypes and the sequences from the literature, these assignments were almost unequivocal. However, as many new sequences have been encountered in the typing exercise mentioned above (C. Tregaskes, R. Martin, H. Afrache and J. Kaufman, unpublished), some clades have grown enormously, so that sequences within a clade can differ by more positions than sequences between clades. Apparently this has also been a problem for the human MHC, for which the first allele groups were easily assigned based on serology, but wide-ranging sequencing led to enormous variation throughout the sequences, so that simple sequence comparisons began to be insufficient to make meaningful assignments (Robinson et al 2017). Among the alternative possibilities would be a classification based on peptide-binding, if it can be related reliably to particular positions in the sequence, such as is attempted with the concept of supertypes (Greenbaum et al 2011; Sidney et al 2008).

In conclusion, the implementation of an IPD-MHC database for chicken MHC sequences has forced the consideration of a new nomenclature system based on gene sequences rather than on haplotypes, which is proposed in this review.

However, the discovery of many hundreds of new alleles in many new haplotypes has highlighted difficulties in this new naming process, the solutions to which are still under consideration. Therefore, the new names in this review should only be considered provisional at best, and may be replaced entirely in the future. Overall, progress is being made, but problems have arisen. With the advent of easier methods for sequencing larger stretches of DNA, the next few years should see more complete sequences of the chicken MHC and hopefully the way forward will become clear.

Acknowledgements. We thank the members (particularly Prof. Steve Marsh, Prof. Ronald Bontrop, Dr. Keith Ballingall and Dr. John Hammond) of the ISAG/IUIS-VIC committee for comparative MHC nomenclature (also serving as the IPD steering committee) for much guidance about nomenclature, and Prof. Mike Ratcliffe and Ms. Rebecca Martin for very useful discussions and for critical reading of the manuscript. This work was funded by a Wellcome Trust Investigator Award (110106/Z/15/Z).

References

Abplanalp H, Hagger C, Briles R (1981) Genetic variation of blood groups in inbred lines of Leghorns, derived from a common base population. J Hered 72:224-226. PubMed PMID: 7276533.

Abraham JP, Barker DJ, Robinson J, Maccari G, Marsh SGE (2018) The IPD Databases: Cataloguing and Understanding Allele Variants. Methods Mol Biol 1802:31-48. doi: 10.1007/978-1-4939-8546-3_3. PubMed PMID: 29858800.

Afanassieff M, Goto RM, Ha J, Sherman MA, Zhong L, Auffray C, Coudert F, Zoorob R, Miller MM (2001) At least one class I gene in restriction fragment pattern-Y (Rfp-Y), the second MHC gene cluster in the chicken, is transcribed, polymorphic, and shows divergent specialization in antigen binding region. J Immunol 166:3324-3333. PubMed PMID: 11207288.

Atkinson D, Shaw I, Jacob J, Kaufman J: DM gene polymorphisms: co-evolution or coincidence? In Proceedings of the Avian Immunology Research Group, 7-10 October 2000, Ithaca NY. Edited by KA Schat; 2001, 163-165.

Bacon LD, Kite JH Jr, Rose NR (1973) Immunogenetic detection of B locus genotypes in chickens with autoimmune thyroiditis. Transplantation 16:591-598. PubMed PMID: 4585383.

Ballingall KT, Bontrop RE, Ellis SA, Grimholt U, Hammond JA, Ho CS, Kaufman J, Kennedy LJ, Maccari G, Miller D, Robinson J, Marsh SGE (2018) Comparative MHC nomenclature: report from the ISAG/IUIS-VIC committee 2018. Immunogenetics 70:625-632. doi: 10.1007/s00251-018-1073-3. PubMed PMID: 30039257.

Baltimore D (1981) Gene conversion: some implications for immunoglobulin genes. Cell 24:592-594. PMID: 7018693 doi: 10.1016/0092-8674(81)90082-9.

Bloom SE, Bacon LD (1985) Linkage of the major histocompatibility (B) complex and the nucleolar organizer in the chicken. Assignment to a microchromosome. J Hered 76:146-154. PubMed PMID: 3998437.

Briles WE, Briles RW, Taffs RE, Stone HA (1983) Resistance to a malignant lymphoma in chickens is mapped to subregion of major histocompatibility (B) complex. Science 219:977-979. PubMed PMID: 6823560.

Briles WE, Bumstead N, Ewert DL, Gilmour DG, Gogusev J, Hala K, Koch C, Longenecker BM, Nordskog AW, Pink JR, Schierman LW, Simonsen M, Toivanen A, Toivanen P, Vainio O, Wick G (1982) Nomenclature for chicken major histocompatibility (B) complex. Immunogenetics 15:441-447. PubMed PMID: 7106862.

Briles WE, McGibbon WH, Irwin MR. On multiple alleles effecting cellular antigens in the chicken (1950) Genetics 35:633-652. PubMed PMID: 14793708; PubMed Central PMCID: PMC1224328.

Briles WE, Stone HA, Cole RK. Marek's disease: effects of B histocompatibility alloalleles in resistant and susceptible chicken lines (1977) Science 195:193-195. PubMed PMID: 831269.

Chappell P, Meziane el K, Harrison M, Magiera Å, Hermann C, Mears L, Wrobel AG, Durant C, Nielsen LL, Buus S, Ternette N, Mwangi W, Butter C, Nair V, Ahyee T, Duggleby R, Madrigal A, Roversi P, Lea SM, Kaufman J (2015) Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. Elife 4:e05345. doi:10.7554/eLife.05345. PubMed PMID: 25860507; PubMed Central PMCID: PMC4420994.

Chattaway J, Ramirez-Valdez RA, Chappell PE, Caesar JJ, Lea SM, Kaufman J (2016) Different modes of variation for each BG lineage suggest different functions. Open Biol 6: 160188. doi: 10.1098/rsob.160188. PubMed PMID: 27628321; PubMed Central PMCID: PMC5043582.

Chazara O, Tixier-Boichard M, Morin V, Zoorob R, Bed'hom B (2011) Organisation and diversity of the class II DM region of the chicken MHC. Mol Immunol 48:1263-1271. doi: 10.1016/j.molimm.2011.03.009. PubMed PMID: 21481938.

Chen F, Pan L, Chao W, Dai Y, Yu W (2012) Character of chicken polymorphic major histocompatibility complex class II alleles of 3 Chinese local breeds. Poult Sci 91:1097-1104. doi: 10.3382/ps.2011-02007. PubMed PMID: 22499866.

Ewald SJ, Livant EJ (2004) Distinctive polymorphism of chicken B-FI (major histocompatibility complex class I) molecules. Poult Sci 83:600-605. PubMed PMID: 15109057.

Fulton JE, Thacker EL, Bacon LD, Hunt HD (1995) Functional analysis of avian class I (BFIV) glycoproteins by epitope tagging and mutagenesis in vitro. Eur J Immunol 25:2069-2076. PubMed PMID: 7621880.

Gebriel GM, Nordskog AW (1983) Genetic linkage of subgroup C Rous sarcoma virus-induced tumour expression in chickens to the IR-GAT locus of the B complex. J Immunogenet 10:231-235. PubMed PMID: 6308101.

Gilmour DG (1959) Segregation of Genes Determining Red Cell Antigens at High Levels of Inbreeding in Chickens. Genetics 44:14-33. PubMed PMID: 17247808; PubMed Central PMCID: PMC1209932.

Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A (2011) Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics 63:325-335. doi: 10.1007/s00251-011-0513-0. PubMed PMID: 21305276; PubMed Central PMCID: PMC3626422.

Guillemot F, Billault A, Pourquie O, Behar G, Chausse AM, Zoorob R, Kreibich G, Auffray C (1988) A molecular map of the chicken major histocompatibility complex: the class II beta genes are closely linked to the class I genes and the nucleolar organizer. EMBO J 7:2775-2785. PubMed PMID: 3141149; PubMed Central PMCID: PMC457068.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98.

Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, Nishibori M,Inoko H, Hanzawa K, Shiina T (2008) Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. J Immunol 181:3393-3399. PubMed PMID: 18714011; PubMed Central PMCID: PMC2657362.

Hunt HD, Fulton JE (1998) Analysis of polymorphisms in the major expressed class I locus (B-FIV) of the chicken. Immunogenetics 47:456-467. PubMed PMID: 9553152.

Hunt HD, Pharr GT, Bacon LD (1994) Molecular analysis reveals MHC class I intra-locus recombination in the chicken. Immunogenetics 40:370-375. PubMed PMID: 7927541.

Jacob JP, Milne S, Beck S, Kaufman J (2000) The major and a minor class II beta-chain (B-LB) gene flank the Tapasin gene in the B-F /B-L region of the chicken major histocompatibility complex. Immunogenetics 51:138-147. PubMed PMID: 10663576.

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275-282.

Juul-Madsen HR, Dalgaard TS, Afanassieff M (2000) Molecular characterization of major and minor MHC class I and II genes in B21-like haplotypes in chickens. Anim Genet 31:252-261. PubMed PMID: 11086534.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066.

Kaufman, J. 2013. The avian MHC. In *Avian Immunology*, 2nd ed., K. A. Schat, P. Kaiser, and B. Kaspers,eds. Academic Press. p. 149-167.

Kaufman J, Andersen R, Avila D, Engberg J, Lambris J, Salomonsen J, Welinder K, Skjodt K (1992) Different features of the MHC class I heterodimer have evolved at different rates. Chicken B-F and beta 2-microglobulin sequences reveal invariant surface residues. J Immunol 148:1532-1546. PubMed PMID: 1538136.

Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923-925. PubMed PMID: 10553909.

Kaufman J, Volk H, Wallny HJ (1995) A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. Immunol Rev 143:63-88. PubMed PMID: 7558083.

Kim T, Hunt HD, Parcells MS, van Santen V, Ewald SJ (2018) Two class I genes of the chicken MHC have different functions: BF1 is recognized by NK cells while BF2 is recognized by CTLs. Immunogenetics 70:599-611. doi: 10.1007/s00251-018-1066-2. PubMed PMID: 29947944.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molec Biol Evol 33:1870-1874.

Li L, Johnson LW, Ewald SJ (1997) Molecular characterization of major histocompatibility complex (B) haplotypes in broiler chickens. Anim Genet 28:258-267. PubMed PMID: 9345722.

Li L, Johnson LW, Livant EJ, Ewald SJ (1999) The MHC of a broiler chicken line: serology, B-G genotypes, and B-F/B-LB sequences. Immunogenetics 49:215-224. PubMed PMID: 9914335.

Lima-Rosa CA, Canal CW, Streck AF, Freitas LB, Delgado-Canedo A, Bonatto SL, Salzano FM (2004) B-F DNA sequence variability in Brazilian (blue-egg Caipira) chickens. Anim Genet 35:278-284. PubMed PMID: 15265066.

Liu W, Miller MM, Lamont SJ (2002) Association of MHC class I and class II gene polymorphisms with vaccine or challenge response to Salmonella enteritidis in young chicks. Immunogenetics 54:582-590. PubMed PMID: 12439621.

Livant EJ, Brigati JR, Ewald SJ (2004) Diversity and locus specificity of chicken MHC B class I sequences. Anim Genet 35:18-27. PubMed PMID: 14731225.

Livant EJ, Zheng D, Johnson LW, Shi W, Ewald SJ (2001) Three new MHC haplotypes in broiler breeder chickens. Anim Genet 32:123-131. PubMed PMID: 11493260.

Lundqvist ML, Middleton DL, Hazard S, Warr GW (2001) The immunoglobulin heavy chain locus of the duck. Genomic organization and expression of D, J, and C region genes. J Biol Chem 276: 46729-46736.

Maccari G, Robinson J, Bontrop RE, Otting N, de Groot NG, Ho CS, Ballingall KT, Marsh SGE, Hammond JA (2018) IPD-MHC: nomenclature requirements for the non-human major histocompatibility complex in the next-generation sequencing era. Immunogenetics 70:619-623. doi: 10.1007/s00251-018-1072-4. PubMed PMID: 30027299; PubMed Central PMCID: PMC6182402.

Marsh SG, Bodmer JG (1991) HLA class II nucleotide sequences, 1991. Immunogenetics 33:321-334. PubMed PMID: 1904836.

McCormack, WT and Thompson CB (1990) Chicken IgL variable region gene conversions display pseudogene donor preference and 5' to 3' polarity. Genes Dev 4: 548-558.

Miller MM, Taylor RL Jr (2016) Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. Poult Sci 95:375-392. doi: 10.3382/ps/pev379. PubMed PMID: 26740135; PubMed Central PMCID: PMC4988538.

Miller MM, Bacon LD, Hala K, Hunt HD, Ewald SJ, Kaufman J, Zoorob R, Briles WE (2004) 2004 Nomenclature for the chicken major histocompatibility (B and Y) complex. Immunogenetics 56:261-279. PubMed PMID: 15257423.

Miller MM, Goto R, Bernot A, Zoorob R, Auffray C, Bumstead N, Briles WE (1994) Two Mhc class I and two Mhc class II genes map to the chicken Rfp-Y system outside the B complex. Proc Natl Acad Sci U S A 91:4397-4401. PubMed PMID: 7910407; PubMed Central PMCID: PMC43792.

Parker A, Kaufman J (2017) What chickens might tell us about the MHC class II system. Curr Opin Immunol 46:23-29. doi: 10.1016/j.coi.2017.03.013. PubMed PMID: 28433952.

Pharr GT, Dodgson JB, Hunt HD, Bacon LD (1998) Class II MHC cDNAs in 15I5 B-congenic chickens. Immunogenetics 47:350-354. PubMed PMID: 9510552.

Plachy J, Pink JR, Hala K (1992) Biology of the chicken MHC (B complex). Crit Rev Immunol 12:47-79. PubMed PMID: 1358107.

Potts ND (2016) Haplotype diversity and stability in the chicken major histocompatibility complex. Dissertation, University of Cambridge.

Riegert P, Andersen R, Bumstead N, Dohring C, Dominguez-Steglich M, Engberg J, Salomonsen J, Schmid M, Schwager J, Skjodt K, Kaufman J (1996) The chicken beta2-microglobulin gene is located on a non-major histocompatibility complex microchromosome: a small, G+C-rich gene with X and Y boxes in the promoter. Proc Natl Acad Sci U S A 93:1243-1248. PubMed PMID: 8577748; PubMed Central PMCID: PMC40064.

Robinson J, Guethlein LA, Cereb N, Yang SY, Norman PJ, Marsh SGE, Parham P (2017) Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. PLoS Genet 13:e1006862. doi: 10.1371/journal.pgen.1006862. PubMed PMID: 28650991; PubMed Central PMCID: PMC5507469.

Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG (2015) The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res 43:D423-431. doi: 10.1093/nar/gku1161. PubMed PMID: 25414341; PubMed Central PMCID: PMC4383959.

Robinson J, Malik A, Parham P, Bodmer JG, Marsh SG (2000) IMGT/HLA database—a sequence database for the human major histocompatibility complex. Tissue Antigens 55:280-287. PubMed PMID: 10777106.

Rogers SL, Kaufman J (2008) High allelic polymorphism, moderate sequence diversity and diversifying selection for B-NK but not B-lec, the pair of lectin-like receptor genes in the chicken MHC. Immunogenetics 60:461-475. doi: 10.1007/s00251-008-0307-1. PubMed PMID: 18574582.

Salomonsen J, Chattaway JA, Chan AC, Parker A, Huguet S, Marston DA, Rogers SL, Wu Z, Smith AL, Staines K, Butter C, Riegert P, Vainio O, Nielsen L, Kaspers B, Griffin DK, Yang F, Zoorob R, Guillemot F, Auffray C, Beck S, Skjodt K, Kaufman J (2014) Sequence of a complete chicken BG haplotype shows dynamic expansion and contraction of two gene lineages with particular expression patterns. PLoS Genet 10:e1004417. doi: 10.1371/journal.pgen.1004417. PubMed PMID: 24901252; PubMed Central PMCID: PMC4046983.

Salomonsen J, Marston D, Avila D, Bumstead N, Johansson B, Juul-Madsen H, Olesen GD, Riegert P, Skjodt K, Vainio O, Wiles MV, Kaufman J (2003) The properties of the single chicken MHC classical class II alpha chain (B-LA) gene indicate an ancient origin for the DR/E-like isotype of class II molecules. Immunogenetics 55:605-614. PubMed PMID: 14608490.

Sayegh CE, Drury G, Ratcliffe MJH (1999) Efficient antibody diversification by gene conversion in vivo in the absence of selection for V(D)J-encoded determinants. EMBO J 18:6319-6328.

Schierman LW, Nordskog AW (1961) Relationship of blood type to histocompatibility in chickens. Science 134:1008-1009. PubMed PMID: 13747603.

Schierman LW, Nordskog AW (1963) Influence of the B blood group-histocompatibility locus in chickens on a graft-versus-host reaction. Nature 197:511-512. PubMed PMID: 13991668.

Shaw I, Powell TJ, Marston DA, Baker K, van Hateren A, Riegert P, Wiles MV, Milne S, Beck S, Kaufman J (2007) Different evolutionary histories of the two classical class I genes BF1 and BF2 illustrate drift and selection within the stable MHC haplotypes of chickens. J Immunol 178:5744-5752. PubMed PMID: 17442958.

Shiina T, Briles WE, Goto RM, Hosomichi K, Yanagiya K, Shimizu S, Inoko H, Miller MM (2007) Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. J Immunol 178:7162-7172. PubMed PMID: 17513765.

Sidney J, Peters B, Frahm N, Brander C, Sette A (2008) HLA class I supertypes: a revised and updated classification. BMC Immunol 9:1. doi: 10.1186/1471-2172-9-1. PubMed PMID: 18211710; PubMed Central PMCID: PMC2245908.

Simonsen M, Crone M, Koch C, Hala K (1982) The MHC haplotypes of the chicken. Immunogenetics 16:513-532. PubMed PMID: 6763913.

Suyama Y, Munkres KD, Woodward VW (1959) Genetic analyses of the pyr-3 locus of Neurospora crassa: the bearing of recombination and gene conversion upon intraallelic linearity. Genetica 30:293-311. PubMed PMID: 13835862.

Suzuki K, Kobayashi E, Yamashita H, Uenishi H, Churkina I, Plastow G, Hamasima N, Mitsuhashi T (2012) Structural analysis of MHC alleles in an RSV tumour regression chicken using a BAC library. Anim Genet 43:348-351. doi: 10.1111/j.1365-2052.2011.02247.x. PubMed PMID: 22486511.

Thacker EL, Fulton JE, Hunt HD (1995) In vitro analysis of a primary, major histocompatibility complex (MHC)-restricted, cytotoxic T-lymphocyte response to avian leukosis virus (ALV), using target cells expressing MHC class I cDNA inserted into a recombinant ALV vector. J Virol 69:6439-6444. PubMed PMID: 7666545; PubMed Central PMCID: PMC189544.

van Hateren A, Carter R, Bailey A, Kontouli N, Williams AP, Kaufman J, Elliott T (2013) A mechanistic basis for the co-evolution of chicken tapasin and major histocompatibility complex class I (MHC I) proteins. J Biol Chem 288:32797-32808. doi: 10.1074/jbc.M113.474031. PubMed PMID: 24078633; PubMed Central PMCID: PMC3820913.

Vilhelmova M, Miggiano VC, Pink JR, Hala K, Hartmanova J (1977) Analysis of the alloimmune properties of a recombinant genotype in the major histocompatibility complex of the chicken. Eur J Immunol 7:674-679. PubMed PMID: 73463.

Walker BA, Hunt LG, Sowa AK, Skjodt K, Gobel TW, Lehner PJ, Kaufman J (2011) The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. Proc Natl Acad Sci U S A 108:8396-8401. doi: 10.1073/pnas.1019496108. PubMed PMID: 21536896; PubMed Central PMCID: PMC3100931.

Walker BA, van Hateren A, Milne S, Beck S, Kaufman J (2005) Chicken TAP genes differ from their human orthologues in locus organisation, size, sequence features and polymorphism. Immunogenetics 57:232-247. PubMed PMID: 15900495.

Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, Skjodt K, Vainio O, Vilbois F, Wiles MV, Kaufman J (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci U S A 103:1434-1439. PubMed PMID: 16432226; PubMed Central PMCID: PMC1360531.

Wittzell H, Bernot A, Auffray C, Zoorob R (1999) Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. Mol Biol Evol 16:479-490. PubMed PMID: 10331274.

Wolf H, Hala K, Boyd RL, Wick G (1984) MHC- and non-MHC-encoded surface antigens of chicken lymphoid cells and erythrocytes recognized by polyclonal xeno-, allo- and monoclonal antibodies. Eur J Immunol 14:831-839. PubMed PMID: 6479208.

Worley K, Gillingham M, Jensen P, Kennedy LJ, Pizzari T, Kaufman J, Richardson DS (2008) Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. Immunogenetics 60:233-247. doi: 10.1007/s00251-008-0288-0. PubMed PMID: 18389232.

Zemmour J, Parham P (1991) HLA class I nucleotide sequences, 1991. Immunogenetics 33:310-320. PubMed PMID: 2050388.

Zhao Y, Rabbani H, Shimizu A, Hammarström L (2000) Mapping of the chicken immunoglobulin heavy-chain constant region gene locus reveals an inverted alpha gene upstream of a condensed upsilon gene. Immunology 101:348-353.

Zoorob R, Bernot A, Renoir DM, Choukri F, Auffray C (1993) Chicken major histocompatibility complex class II B genes: analysis of interallelic and interlocus sequence variance. Eur J Immunol 23:1139-1145. PubMed PMID: 8477808.

Figure legends

Fig. 1. Organization of regions on chicken chromosome 16, as currently understood. a. Depiction of chromosome 16, based on analysis by FISH, radiation hybrids, genetics, southern blotting and sequencing. b, B locus; GC, G+C rich region; Y, Rfp-Y region; NOR, nucleolar organiser region; BLA, class II A gene; fB, factor B gene. Double-headed arrows indicate recombination frequencies for between B and BLA, fB and Rfp-Y, and B and Rfp-Y. B. Region of the B locus currently sequenced, including the BF-BL region, the TRIM region and the BG region. Genes are represented by boxes. Rising and falling stripes indicate genes of the classical class I and class II presentation system, respectively; stippled indicate class III region genes; black indicates lectin-like genes and pseudogenes; horizontal stripes indicate TRIM family genes; vertical stripes indicate BG genes. Names of genes above indicate transcription from left to right, below indicate transcription from right to left. References to support sequence data and identifications in Kaufman 2013, from which this figure and figure legend are taken (with permission), except for the BG region (Salomonsen et al 2014).

Fig. 2. Phylogenetic trees for amino acid sequences of MHC peptide-binding domains from standard haplotypes. a. α1 and α2 domains of BF sequences (with the first and last seven amino acids removed, corresponding to primers and other reasons for different lengths of sequence); b. β1 domains of BLB sequences (with the first two amino acids removed, corresponding to primers and other reasons for different lengths of sequence). Neighbour joining (NJ) trees were created by MEGA7 (Kumar et al 2016) using the Jones-Taylor-Thornton (JTT) matrix-based method (Jones et al 1992) using sequences from the GenBank accession numbers on the tree and with human sequences as outgroups. Genetic distances are indicated with bars; red numbers are bootstrap values (percentages) for those nodes that reach significance from 500 replications; names at the tips are the gene name, followed by the GenBank accession number, followed by the haplotype. Allele groups for BF1 and BLB1 (or BF2 and BLB2) are named, either in green (or blue) for single sequences or in black surrounded by green (or blue) background for clades with more than one sequence; the coloured background for clades with sequences from more than one haplotype are surrounded by a black line.

Fig. 3. Distance matrices for amino acid sequences of MHC peptide-binding domains from standard haplotypes, with the α1 and α2 domains of a. BF1 versus BF1 alleles, b. BF2 versus BF2 alleles, c. BF1 versus BF2 alleles; d. BLB1 versus BLB1 alleles, e. BLB2 versus BLB2 alleles, f. BLB1 versus BLB2 alleles. The sequences used are the consensus full length domains, without truncation. Alignments were performed using MAFFT on-line (Katoh et al. 2002; https://mafft.cbrc.jp/alignment/server/) and the results were pasted into Bioedit (Hall 1999; https://softfamous.com/bioedit/) on a desktop computer; the command “Sequence difference count Matrix” under “Alignment” was used to generate the distance matrix, which was pasted into Microsoft Excel and then Powerpoint for producing the final figure. Highlights indicate differences for BF (or BLB): green, none; blue, 1 to 4 (1 or 2); yellow, 5 to 8 (3 or 4); ID, comparison between the same sequence.

Fig. 4. Haplotypes with genes as boxes for the standard haplotypes, with haplotype strings of the proposed allele names. Comparisons based on a. full coding sequences (CDS), b. peptide-binding domains/exons. Unique sequences for each allelic series are indicated as white boxes, identical sequences are the same colors, and closely-related variants have the same colors but are striped in different ways for different variants of the same allele group. Note: the boxes for some alleles that are striped as variants considering the whole CDS in Fig. 4a are not striped in Fig. 4b since the there are no differences in the peptide-binding domains/exons. GenBank accession numbers and citations for the sequences are in the legend of Online Resource 2.

Fig. 5. Haplotypes with genes as boxes, with haplotype strings of the proposed allele names (with two fields). Comparisons are for peptide-binding domains/exons for a. standard haplotypes, b. sequences from the scientific literature. Unique sequences for each allelic series are indicated as white boxes, identical sequences in the coding region (CDS) are the same colors, and closely-related variants have the same colors but are striped in different ways for different variants of the same allele group. Note: in comparison to Fig. 4b, the boxes for some standard haplotypes are now colored since the sequences, which were unique when only the standard haplotypes were compared, are now in an allelic group with sequences from the scientific literature. Also, some allele names are groups of closely-related variants which cannot be distinguished by the peptide-binding domains/exons alone. GenBank accession numbers and citations for the sequences are in the legend of Online Resource 2.

Fig. 6. The compact nature of the BF-BL region and the inverted orientation of BLB1 and BLB2 can facilitate exchange between the two genes. Boxes indicate exons of BLB1 and BLB2, with solid colours indicating coding sequence and striped colours indicating untranslated regions. Yellow indicates sequence that is identical (or nearly so) between BLB1 and BLB2 genes; green and blue indicates regions that are specific to BLB1 and BLB2, respectively; a grey box indicates the tapasin gene. The conventional organisation is subject to sequence exchange by a. simple inversion of the whole genes or b. “gene conversion” (equivalent to double reciprocal recombination) of exon 2; red X indicate points of recombination and arrows above the genes indicate transcriptional orientation.

Supplementary figure legends

Online Resource 1. Phylogenetic trees for nucleotide sequences of exons encoding MHC peptide-binding domains from all chicken MHC-like sequences found in GenBank. a. exons 2 and 3 of chicken class I sequences; b. exon 2 of chicken class II B sequences. Sequences were found in two ways from the non-redundant GenBank database [National Center for Biotechnology Information (NCBI); www.ncbi.nlm.nih.gov/nuccore/] accessed in January 2017: by keywords (chicken AND MHC I AND gallus gallus, gallus gallus AND class I AND chromosome 16, g.gallus AND MHC class I, chicken AND MHC I AND gallus gallus AND chromosome 16), and by BLAST search using default parameters for nucleotide sequences (except using 1000 hits) and using AB426141 for BF and AB426141 for BLB as query sequences. Neighbour joining (NJ) trees were created by MEGA7 [Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molec Biol Evol 33:1870-1874] using Tamura-Nei method [Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molec Biol Evol 4:406-425]. Genetic distances are indicated with bars; red numbers are bootstrap values (percentages) for those nodes that reach significance from 500 replications; names at the tips are the GenBank accession number (followed by the accession number for the CDS protein sequence for some entries).

Online Resource 2. Tables of BF and BLB names with accession numbers and citations for standard haplotypes and literature alleles; bold entries are for sequences from the standard haplotypes. Citations:

Chen F, Pan L, Chao W, Dai Y, Yu W (2012) Character of chicken polymorphic major histocompatibility complex class II alleles of 3 Chinese local breeds. Poult Sci 91:1097-1104. doi: 10.3382/ps.2011-02007. PubMed PMID: 22499866.

Dalgaard TS, Vitved L, SkjÃ¸dt K, Thomsen B, Labouriau R, Jensen KH, Juul-Madsen HR (2005) Molecular characterization of major histocompatibility complex class I (B-F) mRNA variants from chickens differing in resistance to Marek's disease. Scand J Immunol 62:259-270. PubMed PMID: 16179013

Fulton JE, Thacker EL, Bacon LD, Hunt HD (1995) Functional analysis of avian class I (BFIV) glycoproteins by epitope tagging and mutagenesis in vitro. Eur J Immunol 25:2069-2076. PubMed PMID: 7621880.

Goto RM, Wang Y, Taylor RL Jr, Wakenell PS, Hosomichi K, Shiina T, Blackmore CS, Briles WE, Miller MM (2009) BG1 has a major role in MHC-linked resistance to malignant lymphoma in the chicken. Proc Natl Acad Sci U S A 106:16740-16745. doi: 10.1073/pnas.0906776106.PubMed PMID: 19805366; PubMedCentral PMCID: PMC2757851.

Guillemot F, Billault A, Pourquié O, Béhar G, Chaussé AM, Zoorob R, Kreibich G, Auffray C. (1998) A molecular map of the chicken major histocompatibility complex: the class II beta genes are closely linked to the class I genes and the nucleolar organizer. EMBO J 7:2775-2785. PubMed PMID: 3141149; PubMed CentralPMCID: PMC457068.

Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, Nishibori M,Inoko H, Hanzawa K, Shiina T (2008) Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. J Immunol 181:3393-3399. PubMed PMID: 18714011; PubMed Central PMCID: PMC2657362.

Hunt HD, Fulton JE (1998) Analysis of polymorphisms in the major expressed class I locus (B-FIV) of the chicken. Immunogenetics 47:456-467. PubMed PMID: 9553152.

Hunt HD, Pharr GT, Bacon LD (1994) Molecular analysis reveals MHC class I intra-locus recombination in the chicken. Immunogenetics 40:370-375. PubMed PMID: 7927541.

Jacob JP, Milne S, Beck S, Kaufman J (2000) The major and a minor class II beta-chain (B-LB) gene flank the Tapasin gene in the B-F /B-L region of the chicken major histocompatibility complex. Immunogenetics 51:138-147. PubMed PMID: 10663576.

Juul-Madsen HR, Dalgaard TS, Afanassieff M (2000) Molecular characterization of major and minor MHC class I and II genes in B21-like haplotypes in chickens. Anim Genet 31:252-261. PubMed PMID: 11086534.

Kaufman J, Andersen R, Avila D, Engberg J, Lambris J, Salomonsen J, Welinder K, Skjodt K (1992) Different features of the MHC class I heterodimer have evolved at different rates. Chicken B-F and beta 2-microglobulin sequences reveal invariant surface residues. J Immunol 148:1532-1546. PubMed PMID: 1538136.

Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923-925. PubMed PMID: 10553909.

Kroemer G, Zoorob R, Auffray C (1990) Structure and expression of a chicken MHC class I gene. Immunogenetics 31:405-409. PubMed PMID: 2370087.

Li L, Johnson LW, Ewald SJ (1997) Molecular characterization of major histocompatibility complex (B) haplotypes in broiler chickens. Anim Genet 28:258-267. PubMed PMID: 9345722.

Li L, Johnson LW, Livant EJ, Ewald SJ (1999) The MHC of a broiler chicken line: serology, B-G genotypes, and B-F/B-LB sequences. Immunogenetics 49:215-224. PubMed PMID: 9914335.

Lima-Rosa CA, Canal CW, Streck AF, Freitas LB, Delgado-Canedo A, Bonatto SL, Salzano FM (2004) B-F DNA sequence variability in Brazilian (blue-egg Caipira) chickens. Anim Genet 35:278-284. PubMed PMID: 15265066.

Liu W, Miller MM, Lamont SJ (2002) Association of MHC class I and class II gene polymorphisms with vaccine or challenge response to Salmonella enteritidis in young chicks. Immunogenetics 54:582-590. PubMed PMID: 12439621.

Livant EJ, Zheng D, Johnson LW, Shi W, Ewald SJ (2001) Three new MHC haplotypes in broiler breeder chickens. Anim Genet 32:123-131. PubMed PMID: 11493260.

Livant EJ, Brigati JR, Ewald SJ (2004) Diversity and locus specificity of chicken MHC B class I sequences. Anim Genet 35:18-27. PubMed PMID: 14731225.

Niemiec PK, Read LR, Sharif S. (2006) Synthesis of chicken major histocompatibility complex class II oligomers using a baculovirus expression system. Protein Expr Purif 46:390-400. PubMed PMID: 16236525.

Pharr GT, Bacon LD, Dodgson JB (1993) Analysis of B-L beta-chain gene expression in two chicken cDNA libraries. Immunogenetics 37:381-385. PubMed PMID:8428771.

Pharr GT, Dodgson JB, Hunt HD, Bacon LD (1998) Class II MHC cDNAs in 15I5 B-congenic chickens. Immunogenetics 47:350-354. PubMed PMID: 9510552.

Shaw I, Powell TJ, Marston DA, Baker K, van Hateren A, Riegert P, Wiles MV, Milne S, Beck S, Kaufman J (2007) Different evolutionary histories of the two classical class I genes BF1 and BF2 illustrate drift and selection within the stable MHC haplotypes of chickens. J Immunol 178:5744-5752. PubMed PMID: 17442958.

Shiina T, Briles WE, Goto RM, Hosomichi K, Yanagiya K, Shimizu S, Inoko H, Miller MM (2007) Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. J Immunol 178:7162-7172. PubMed PMID: 17513765.

Suzuki K, Kobayashi E, Yamashita H, Uenishi H, Churkina I, Plastow G, Hamasima N, Mitsuhashi T (2012) Structural analysis of MHC alleles in an RSV tumour regression chicken using a BAC library. Anim Genet 43:348-351. doi: 10.1111/j.1365-2052.2011.02247.x. PubMed PMID: 22486511.

Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, Skjodt K, Vainio O, Vilbois F, Wiles MV, Kaufman J (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci U S A 103:1434-1439. PubMed PMID: 16432226; PubMed Central PMCID: PMC1360531.

Worley K, Gillingham M, Jensen P, Kennedy LJ, Pizzari T, Kaufman J, Richardson DS (2008) Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. Immunogenetics 60:233-247. doi: 10.1007/s00251-008-0288-0. PubMed PMID: 18389232.

Xu R, Li K, Chen G, Xu H, Qiang B, Li C, Liu B (2007) Characterization of genetic polymorphism of novel MHC B-LB II alleles in Chinese indigenous chickens. J Genet Genomics 34:109-118. PubMed PMID: 17469783.

Zheng D, O'Keefe G, Li L, Johnson LW, Ewald SJ (1999) A PCR method for typing B-L beta II family (class II MHC) alleles in broiler chickens. Anim Genet 30:109-119. PubMed PMID: 10376301.

Zhou H, Lamont SJ (2003) Chicken MHC class I and II gene effects on antibody response kinetics in adult chickens. Immunogenetics 55:133-140. doi: 10.1007/s00251-003-0566-9. PubMed PMID: 12743657.

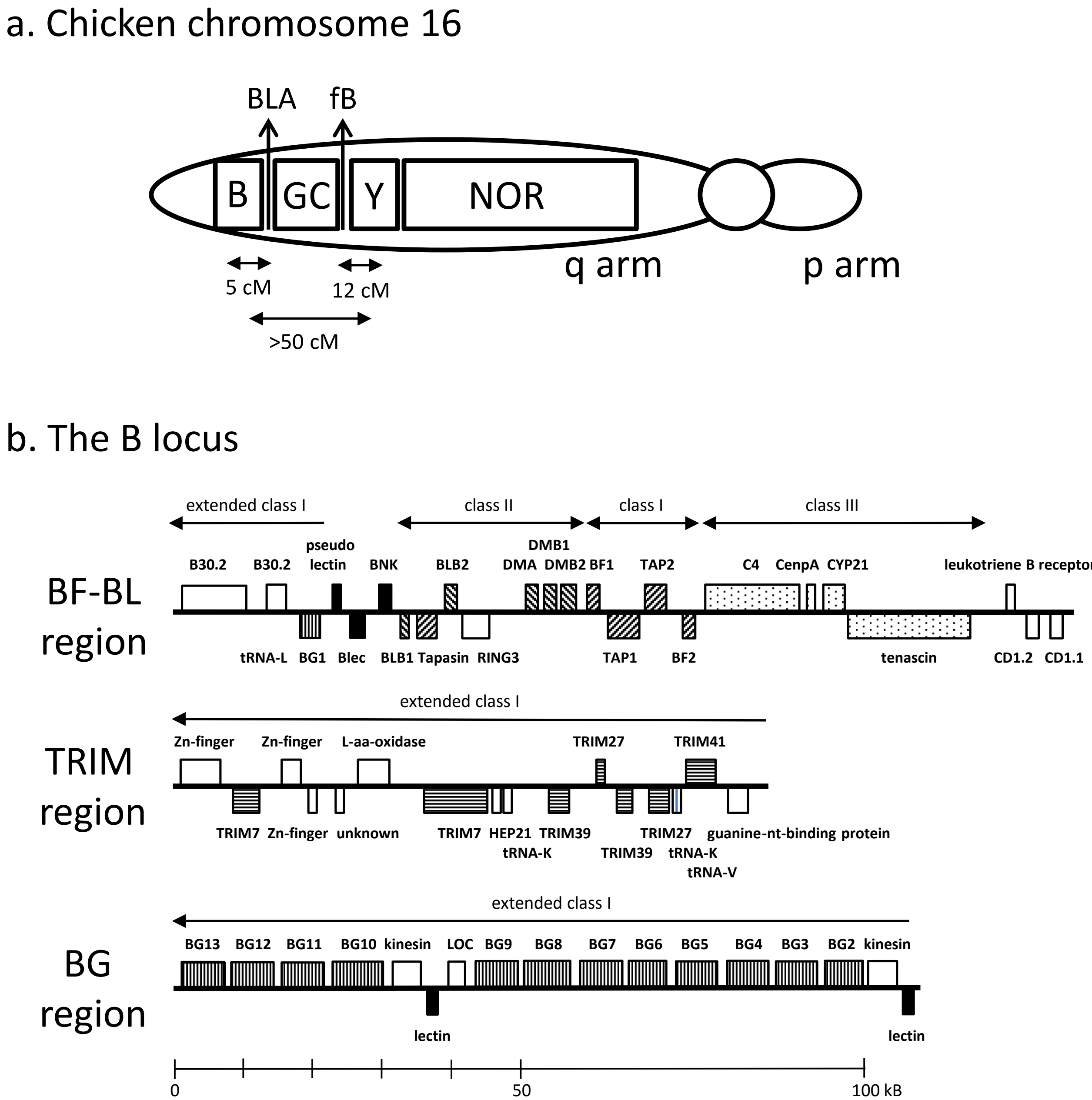
Zoorob R , Béhar G, Kroemer G, Auffray C (1990) Organization of a functional chicken class II B gene. Immunogenetics 31:179-187. PubMed PMID: 1969383.

Online Resource 3. Phylogenetic trees for nucleotide sequences of exons encoding MHC peptide-binding domains from standard haplotypes. a. exons 2 and 3 of BF sequences (with the first 20 nucleotides of exon 2 and last 23 nucleotides of exon 3 removed, corresponding to primers and other reasons for different lengths of sequence); b. exon 2 of BLB sequences (with the first 6-8 nucleotides removed, corresponding to primers and other reasons for different lengths of sequence). Genetic distances are indicated with bars; red numbers are bootstrap values (percentages) for those nodes that reach significance from 500 replications; names at the tips are the gene name, followed by the GenBank accession number, followed by the haplotype. Allele groups for BF1 and BLB1 (or BF2 and BLB2) are named, either in green (or blue) for single sequences or in black surrounded by green (or blue) background for clades with more than one sequence; the coloured background for clades with sequences from more than one haplotype are surrounded by a black line. Human sequences were used as outgroups; sequences for standard haplotypes were taken from the GenBank accession numbers in Online Resource 2; all other details are as in the legend to Online Resource 1.

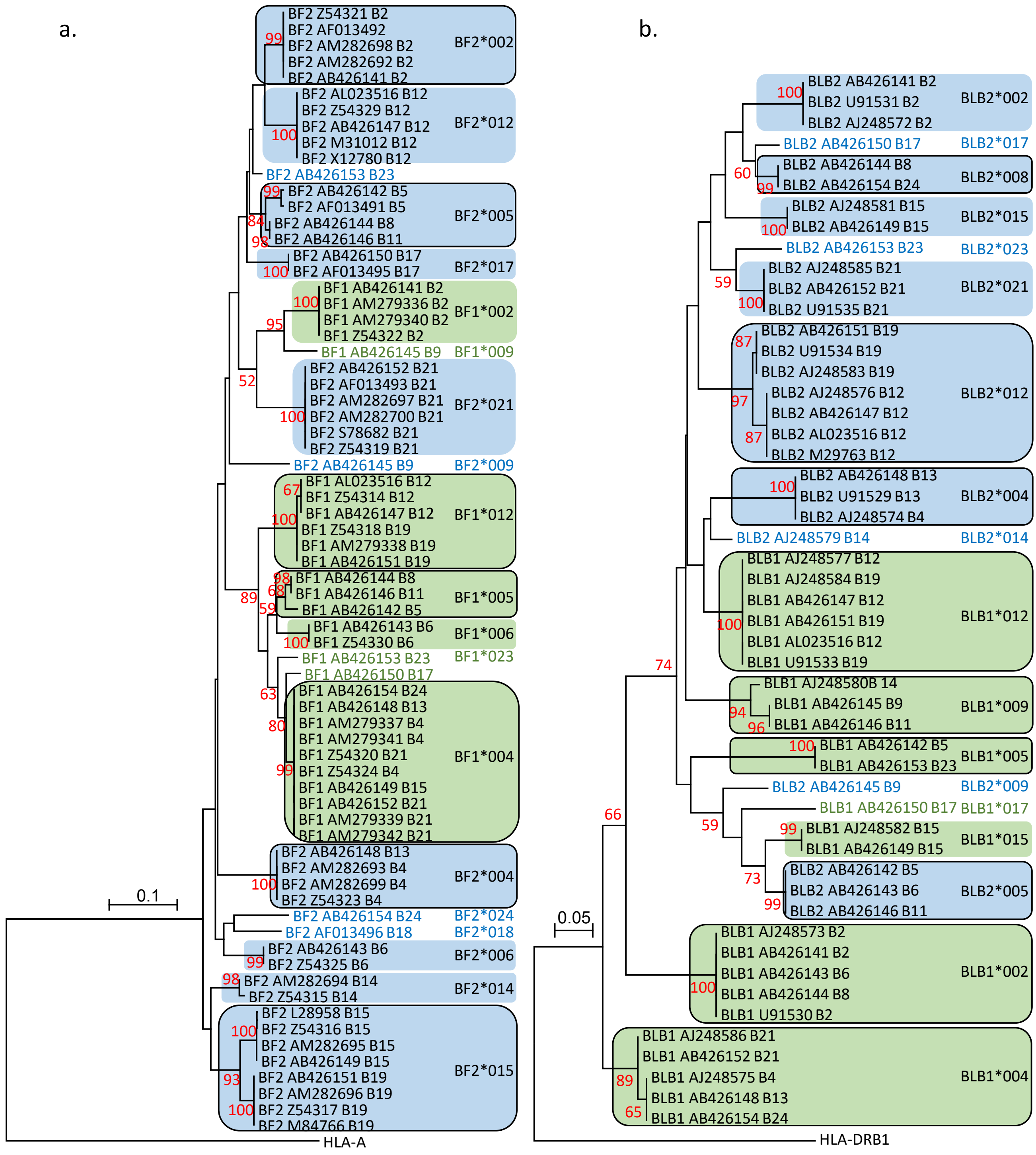
Online resource 4. Distance matrices for nucleotide sequences of exons encoding MHC peptide-binding domains from standard haplotypes. Exons 2 and 3 of a. BF1 versus BF1 alleles, b. BF2 versus BF2 alleles, c. BF1 versus BF2 alleles; d. BLB1 versus BLB1 alleles, e. BLB2 versus BLB2 alleles, f. BLB1 versus BLB2 alleles. Sequences for standard haplotypes were taken from the GenBank accession numbers in Online Resource 2. Alignments were performed using MAFFT on-line [Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066; https://mafft.cbrc.jp/alignment/server/] and the results were pasted into Bioedit [Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98; https://softfamous.com/bioedit/] on a desktop computer; the command "Sequence difference count Matrix" under "Alignment" was used to generate the distance matrix, which was pasted into Microsoft Excel and then Powerpoint for producing the final figure. Highlights indicate amino acid differences for BF (or BLB) from Fig. 3 (for comparison to nucleotide differences in this figure): green, none; blue, 1 to 4 (1 or 2); yellow, 5 to 8 (3 or 4); ID, comparison between the same sequence.

Online Resource 5. Distance matrices for amino acids of whole coding sequences (CDS) from standard haplotypes. a. BF1 versus BF1 alleles, b. BF2 versus BF2 alleles, c. BF1 versus BF2 alleles; d. BLB1 versus BLB1 alleles, e. BLB2 versus BLB2 alleles, f. BLB1 versus BLB2 alleles. All details as in legend to Online Resource 4, except that the names of sequences with indels are highlighted: grey, insertion (one amino acid in BF1, two amino acids in BF2); pink, deletion (five amino acids in BF1, 11 amino acids in BF2); orange, a combination of a one nucleotide deletion and a five nucleotide truncation leading to a frameshift at amino acid 247 in the transmembrane region (BLB1 from the B23 haplotype).

Online Resource 6. Phylogenetic trees for nucleotide sequences of exons encoding MHC peptide-binding domains from standard haplotypes and from sequences in the scientific literature. a. exons 2 and 3 of BF sequences (with the first 20 nucleotides of exon 2 and the last 23 nucleotides of exon 3 removed, corresponding to primers and other reasons for different lengths of sequence); b. exon 2 of BLB sequences (with the first 6-8 nucleotides removed, corresponding to primers and other reasons for different lengths of sequence). Sequences for standard haplotypes and from the scientific literature were taken from the GenBank accession numbers in Online Resource 2; all other details are as in the legends to Online Resources 1 and 3.



a. Chicken chromosome 16
BLA
fB
B
GC
Y
NOR
q arm
p arm
5 cM
12 cM
>50 cM
b. The B locus
extended class I
class II
class I
class III
BF-BL region
B30.2
B30.2
pseudo lectin
BNK
BLB2
DMA
DMB1
DMB2
BF1
TAP2
C4
CenpA
CYP21
leukotriene B receptor
tRNA-L
BG1
Blec
BLB1
Tapasin
RING3
TAP1
BF2
tenascin
CD1.2
CD1.1
TRIM region
extended class I
Zn-finger
Zn-finger
L-aa-oxidase
TRIM27
TRIM41
TRIM7
Zn-finger
unknown
TRIM7
HEP21
tRNA-K
TRIM39
TRIM39
TRIM27
tRNA-K
tRNA-V
guanine-nt-binding protein
BG region
extended class I
BG13
BG12
BG11
BG10
kinesin
LOC
BG9
BG8
BG7
BG6
BG5
BG4
BG3
BG2
kinesin
lectin
lectin
0
50
100 kB


a.
0.1
BF2 Z54321 B2
BF2 AF013492
BF2 AM282698 B2
BF2 AM282692 B2
BF2 AB426141 B2
BF2*002
BF2 AL023516 B12
BF2 Z54329 B12
BF2 AB426147 B12
BF2 M31012 B12
BF2 X12780 B12
BF2*012
BF2 AB426153 B23
BF2 AB426142 B5
BF2 AF013491 B5
BF2 AB426144 B8
BF2 AB426146 B11
BF2*005
BF2 AB426150 B17
BF2 AF013495 B17
BF2*017
BF1 AB426141 B2
BF1 AM279336 B2
BF1 AM279340 B2
BF1 Z54322 B2
BF1*002
BF1 AB426145 B9
BF1*009
BF2 AB426152 B21
BF2 AF013493 B21
BF2 AM282697 B21
BF2 AM282700 B21
BF2 S78682 B21
BF2 Z54319 B21
BF2*021
BF2 AB426145 B9
BF2*009
BF1 AL023516 B12
BF1 Z54314 B12
BF1 AB426147 B12
BF1 Z54318 B19
BF1 AM279338 B19
BF1 AB426151 B19
BF1*012
BF1 AB426144 B8
BF1 AB426146 B11
BF1 AB426142 B5
BF1*005
BF1 AB426143 B6
BF1 Z54330 B6
BF1*006
BF1 AB426153 B23
BF1*023
BF1 AB426150 B17
BF1 AB426154 B24
BF1 AB426148 B13
BF1 AM279337 B4
BF1 AM279341 B4
BF1 Z54320 B21
BF1 Z54324 B4
BF1 AB426149 B15
BF1 AB426152 B21
BF1 AM279339 B21
BF1 AM279342 B21
BF1*004
BF2 AB426148 B13
BF2 AM282693 B4
BF2 AM282699 B4
BF2 Z54323 B4
BF2*004
BF2 AB426154 B24
BF2*024
BF2 AF013496 B18
BF2*018
BF2 AB426143 B6
BF2 Z54325 B6
BF2*006
BF2 AM282694 B14
BF2 Z54315 B14
BF2*014
BF2 L28958 B15
BF2 Z54316 B15
BF2 AM282695 B15
BF2 AB426149 B15
BF2 AB426151 B19
BF2 AM282696 B19
BF2 Z54317 B19
BF2 M84766 B19
BF2*015
HLA-A
b.
0.05
BLB2 AB426141 B2
BLB2 U91531 B2
BLB2 AJ248572 B2
BLB2*002
BLB2 AB426150 B17
BLB2*017
BLB2 AB426144 B8
BLB2 AB426154 B24
BLB2*008
BLB2 AJ248581 B15
BLB2 AB426149 B15
BLB2*015
BLB2 AB426153 B23
BLB2*023
BLB2 AJ248585 B21
BLB2 AB426152 B21
BLB2 U91535 B21
BLB2*021
BLB2 AB426151 B19
BLB2 U91534 B19
BLB2 AJ248583 B19
BLB2 AJ248576 B12
BLB2 AB426147 B12
BLB2 AL023516 B12
BLB2 M29763 B12
BLB2*012
BLB2 AB426148 B13
BLB2 U91529 B13
BLB2 AJ248574 B4
BLB2*004
BLB2 AJ248579 B14
BLB2*014
BLB1 AJ248577 B12
BLB1 AJ248584 B19
BLB1 AB426147 B12
BLB1 AB426151 B19
BLB1 AL023516 B12
BLB1 U91533 B19
BLB1*012
BLB1 AJ248580B 14
BLB1 AB426145 B9
BLB1 AB426146 B11
BLB1*009
BLB1 AB426142 B5
BLB1 AB426153 B23
BLB1*005
BLB2 AB426145 B9
BLB2*009
BLB1 AB426150 B17
BLB1*017
BLB1 AJ248582 B15
BLB1 AB426149 B15
BLB1*015
BLB2 AB426142 B5
BLB2 AB426143 B6
BLB2 AB426146 B11
BLB2*005
BLB1 AJ248573 B2
BLB1 AB426141 B2
BLB1 AB426143 B6
BLB1 AB426144 B8
BLB1 U91530 B2
BLB1*002
BLB1 AJ248586 B21
BLB1 AB426152 B21
BLB1 AJ248575 B4
BLB1 AB426148 B13
BLB1 AB426154 B24
BLB1*004
HLA-DRB1

a.

BF1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B15	B17	B19	B21	B23	B24
B2	ID														
B4	34	ID													
B5	30	13	ID												
B6	29	16	14	ID											
B8	26	11	4	10	ID										
B9	15	35	27	34	31	ID									
B11	26	11	4	10	0	31	ID								
B12	35	19	17	24	21	32	21	ID							
B13	34	0	13	16	11	35	11	19	ID						
B15	34	0	13	16	11	35	11	19	0	ID					
B17	34	5	13	16	11	34	11	20	5	5	ID				
B19	35	18	16	23	20	32	20	1	18	18	19	ID			
B21	34	0	13	16	11	35	11	19	0	0	5	18	ID		
B23	36	9	14	20	15	32	15	16	9	9	11	15	9	ID	
B24	34	0	13	16	11	35	11	19	0	0	5	18	0	9	ID

d.

BLB1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B14	B15	B17	B19	B21	B23	B24
B2	ID															
B4	17	ID														
B5	23	25	ID													
B6	0	17	23	ID												
B8	0	17	23	0	ID											
B9	21	19	21	21	21	ID										
B11	21	19	21	21	21	0	ID									
B12	18	17	20	18	18	15	15	ID								
B13	17	0	25	17	17	19	19	17	ID							
B14	18	19	19	18	18	3	3	14	19	ID						
B15	22	24	20	22	22	17	17	18	24	16	ID					
B17	22	26	18	22	22	20	20	20	26	19	13	ID				
B19	18	17	20	18	18	15	15	0	17	14	18	20	ID			
B21	16	1	24	16	16	20	20	16	1	20	23	25	16	ID		
B23	23	25	0	23	23	21	21	20	25	19	20	18	20	24	ID	
B24	17	0	25	17	17	19	19	17	0	19	24	26	17	1	25	ID

b.

BF2	B2	B4	B5	B6	B8	B9	B11	B12	B13	B14	B15	B17	B18	B19	B21	B23	B24
B2	ID																
B4	25	ID															
B5	13	28	ID														
B6	27	25	26	ID													
B8	11	27	5	26	ID												
B9	24	30	24	28	25	ID											
B11	11	27	5	26	0	25	ID										
B12	11	16	20	25	19	25	19	ID									
B13	25	0	28	25	27	30	27	16	ID								
B14	19	26	19	27	18	24	18	22	26	ID							
B15	22	30	24	26	23	25	23	30	30	15	ID						
B17	16	29	14	28	16	25	16	23	29	17	22	ID					
B18	23	23	23	22	24	28	24	23	23	19	28	22	ID				
B19	20	27	20	24	21	25	21	26	27	17	7	21	28	ID			
B21	20	28	20	30	21	27	21	28	28	23	23	21	29	18	ID		
B23	5	24	9	23	6	24	6	15	24	15	20	11	19	18	20	ID	
B24	24	28	25	22	28	30	28	25	28	30	33	25	22	33	32	24	ID

e.

BLB2	B2	B4	B5	B6	B8	B9	B11	B12	B13	B14	B15	B17	B19	B21	B23	B24
B2	ID															
B4	20	ID														
B5	22	20	ID													
B6	22	20	0	ID												
B8	8	19	21	21	ID											
B9	21	20	9	9	19	ID										
B11	22	20	0	0	21	9	ID									
B12	11	19	19	19	14	16	19	ID								
B13	20	0	20	20	19	20	20	19	ID							
B14	17	11	15	15	15	11	15	13	11	ID						
B15	14	20	22	22	10	21	22	17	20	17	ID					
B17	11	18	19	19	5	17	19	15	18	15	14	ID				
B19	11	19	19	19	12	16	19	2	19	14	15	13	ID			
B21	14	16	18	18	9	16	18	13	16	13	12	10	13	ID		
B23	18	17	20	20	15	18	20	15	17	15	16	12	15	8	ID	
B24	8	19	21	21	0	19	21	14	19	15	10	5	12	9	15	ID

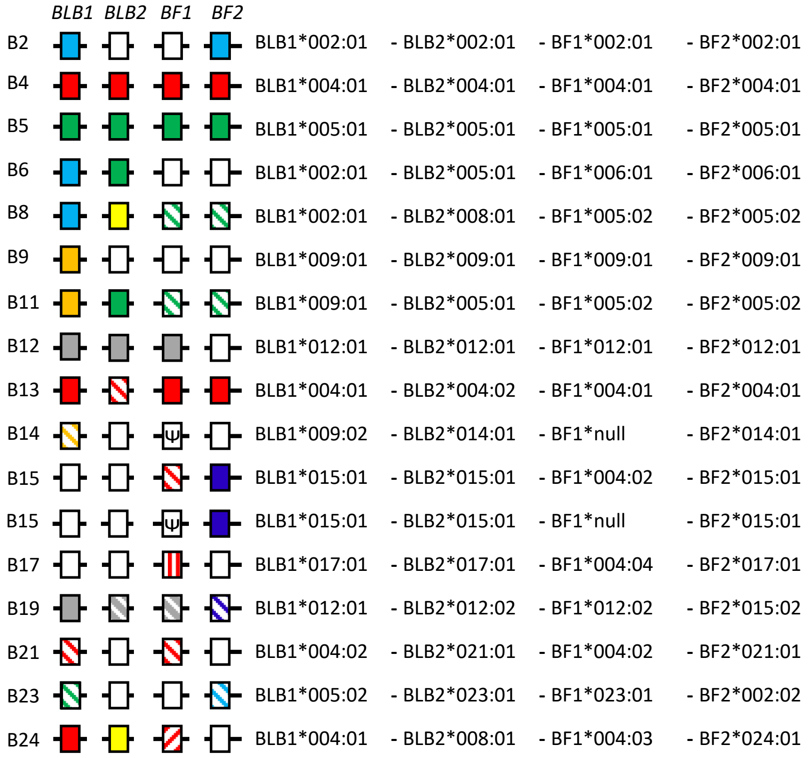
c.

BF2/BF1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B15	B17	B19	B21	B23	B24
B2	29	27	23	31	27	27	27	28	27	27	26	28	27	27	27
B4	27	26	34	31	30	31	30	31	26	26	29	31	26	29	26
B5	28	28	24	29	28	25	28	25	28	28	27	25	28	28	28
B6	25	27	28	20	24	29	24	32	27	27	28	32	27	29	27
B8	27	23	19	30	23	23	23	24	23	23	22	24	23	23	23
B9	31	28	26	32	28	28	28	27	28	28	31	27	28	31	28
B11	27	23	19	30	23	23	23	24	23	23	22	24	23	23	23
B12	33	30	28	33	32	31	32	29	30	30	32	29	30	28	30
B13	27	26	34	31	30	31	30	31	26	26	29	31	26	29	26
B14	32	29	24	29	28	28	28	24	29	29	26	24	29	30	29
B15	26	30	28	31	30	27	30	31	30	30	27	31	30	33	30
B17	33	28	24	28	28	28	28	28	28	28	25	28	28	27	28
B18	33	31	32	29	30	35	30	31	31	31	30	31	31	33	31
B19	26	31	28	31	32	25	32	31	31	31	29	31	31	33	31
B21	26	34	28	30	32	22	32	30	34	34	32	30	34	34	34
B23	27	22	18	27	22	24	22	23	22	22	21	23	22	22	22
B24	33	29	31	26	27	37	27	34	29	29	31	34	29	31	29

f.

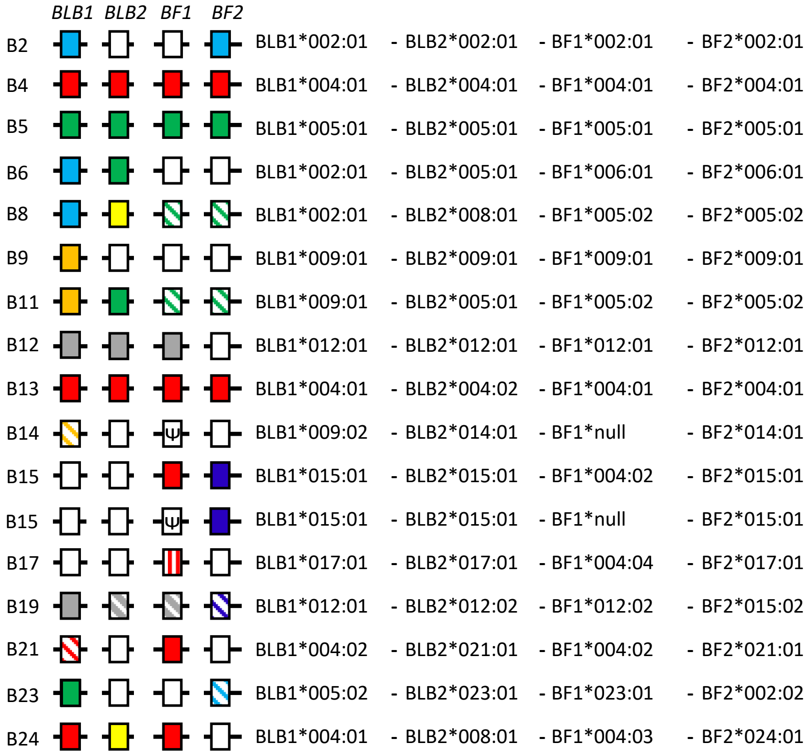
BLB2/BLB1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B14	B15	B17	B19	B21	B23	B24
B2	24	21	25	24	24	21	21	15	21	21	23	23	15	20	25	21
B4	21	23	22	21	21	14	14	13	23	17	21	21	13	22	22	23
B5	25	23	20	25	25	18	18	17	23	17	6	12	17	22	20	23
B6	25	23	20	25	25	18	18	17	23	17	6	12	17	22	20	23
B8	24	20	23	24	24	19	19	13	20	19	23	24	13	19	23	20
B9	21	20	22	21	21	18	18	16	20	17	9	17	16	19	22	20
B11	25	23	20	25	25	18	18	17	23	17	6	12	17	22	20	23
B12	22	18	23	22	22	16	16	13	18	16	22	21	13	17	23	18
B13	21	23	22	21	21	14	14	13	23	17	21	21	13	22	22	23
B14	17	18	21	17	17	12	12	7	18	11	17	18	7	17	21	18
B15	18	20	21	18	18	18	18	14	20	18	22	24	14	19	21	20
B17	22	19	21	22	22	19	19	14	19	17	22	23	14	18	21	19
B19	23	16	21	23	23	16	16	11	16	16	22	21	11	15	21	16
B21	20	21	19	20	20	15	15	14	21	13	19	22	14	20	19	21
B23	20	18	20	20	20	13	13	16	18	11	20	23	16	19	20	18
B24	24	20	23	24	24	19	19	13	20	19	23	24	13	19	23	20

a.

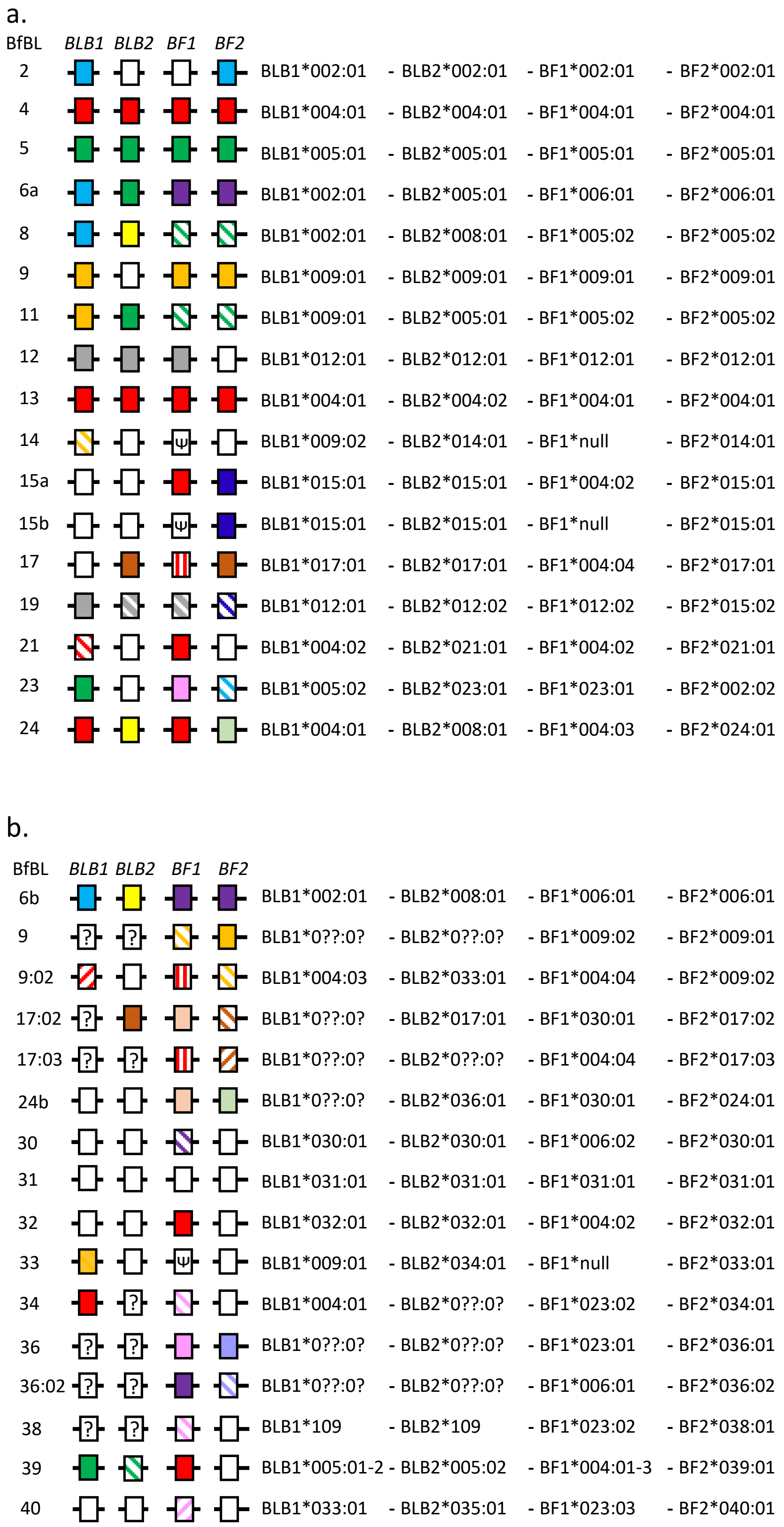


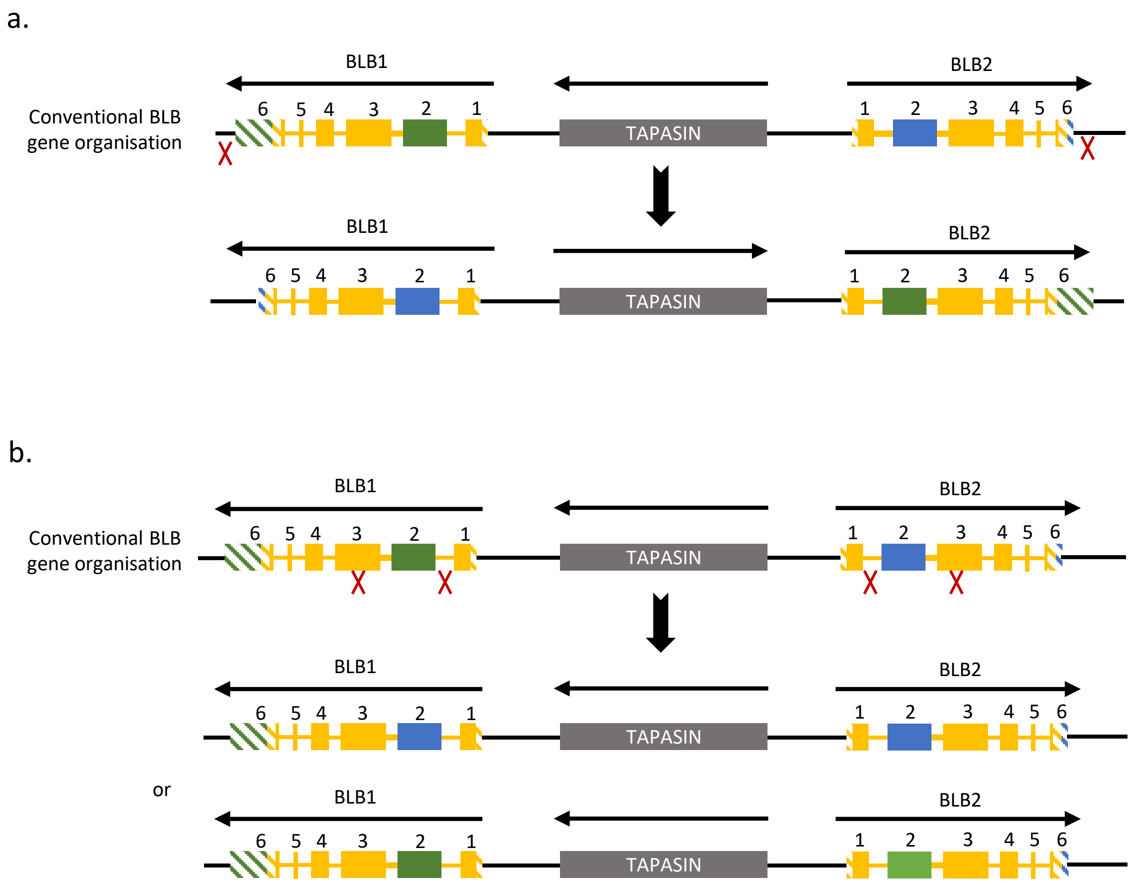
BLB1 BLB2 BF1 BF2
B2 BLB1*002:01 - BLB2*002:01 - BF1*002:01 - BF2*002:01
B4 BLB1*004:01 - BLB2*004:01 - BF1*004:01 - BF2*004:01
B5 BLB1*005:01 - BLB2*005:01 - BF1*005:01 - BF2*005:01
B6 BLB1*002:01 - BLB2*005:01 - BF1*006:01 - BF2*006:01
B8 BLB1*002:01 - BLB2*008:01 - BF1*005:02 - BF2*005:02
B9 BLB1*009:01 - BLB2*009:01 - BF1*009:01 - BF2*009:01
B11 BLB1*009:01 - BLB2*005:01 - BF1*005:02 - BF2*005:02
B12 BLB1*012:01 - BLB2*012:01 - BF1*012:01 - BF2*012:01
B13 BLB1*004:01 - BLB2*004:02 - BF1*004:01 - BF2*004:01
B14 Ψ BLB1*009:02 - BLB2*014:01 - BF1*null - BF2*014:01
B15 BLB1*015:01 - BLB2*015:01 - BF1*004:02 - BF2*015:01
B15 Ψ BLB1*015:01 - BLB2*015:01 - BF1*null - BF2*015:01
B17 BLB1*017:01 - BLB2*017:01 - BF1*004:04 - BF2*017:01
B19 BLB1*012:01 - BLB2*012:02 - BF1*012:02 - BF2*015:02
B21 BLB1*004:02 - BLB2*021:01 - BF1*004:02 - BF2*021:01
B23 BLB1*005:02 - BLB2*023:01 - BF1*023:01 - BF2*002:02
B24 BLB1*004:01 - BLB2*008:01 - BF1*004:03 - BF2*024:01

b.



BLB1 BLB2 BF1 BF2
B2 BLB1*002:01 - BLB2*002:01 - BF1*002:01 - BF2*002:01
B4 BLB1*004:01 - BLB2*004:01 - BF1*004:01 - BF2*004:01
B5 BLB1*005:01 - BLB2*005:01 - BF1*005:01 - BF2*005:01
B6 BLB1*002:01 - BLB2*005:01 - BF1*006:01 - BF2*006:01
B8 BLB1*002:01 - BLB2*008:01 - BF1*005:02 - BF2*005:02
B9 BLB1*009:01 - BLB2*009:01 - BF1*009:01 - BF2*009:01
B11 BLB1*009:01 - BLB2*005:01 - BF1*005:02 - BF2*005:02
B12 BLB1*012:01 - BLB2*012:01 - BF1*012:01 - BF2*012:01
B13 BLB1*004:01 - BLB2*004:02 - BF1*004:01 - BF2*004:01
B14 Ψ BLB1*009:02 - BLB2*014:01 - BF1*null - BF2*014:01
B15 BLB1*015:01 - BLB2*015:01 - BF1*004:02 - BF2*015:01
B15 Ψ BLB1*015:01 - BLB2*015:01 - BF1*null - BF2*015:01
B17 BLB1*017:01 - BLB2*017:01 - BF1*004:04 - BF2*017:01
B19 BLB1*012:01 - BLB2*012:02 - BF1*012:02 - BF2*015:02
B21 BLB1*004:02 - BLB2*021:01 - BF1*004:02 - BF2*021:01
B23 BLB1*005:02 - BLB2*023:01 - BF1*023:01 - BF2*002:02
B24 BLB1*004:01 - BLB2*008:01 - BF1*004:03 - BF2*024:01


a.
BfBL BLB1 BLB2 BF1 BF2
2 BLB1*002:01 - BLB2*002:01 - BF1*002:01 - BF2*002:01
4 BLB1*004:01 - BLB2*004:01 - BF1*004:01 - BF2*004:01
5 BLB1*005:01 - BLB2*005:01 - BF1*005:01 - BF2*005:01
6a BLB1*002:01 - BLB2*005:01 - BF1*006:01 - BF2*006:01
8 BLB1*002:01 - BLB2*008:01 - BF1*005:02 - BF2*005:02
9 BLB1*009:01 - BLB2*009:01 - BF1*009:01 - BF2*009:01
11 BLB1*009:01 - BLB2*005:01 - BF1*005:02 - BF2*005:02
12 BLB1*012:01 - BLB2*012:01 - BF1*012:01 - BF2*012:01
13 BLB1*004:01 - BLB2*004:02 - BF1*004:01 - BF2*004:01
14 BLB1*009:02 - BLB2*014:01 - BF1*null - BF2*014:01
15a BLB1*015:01 - BLB2*015:01 - BF1*004:02 - BF2*015:01
15b BLB1*015:01 - BLB2*015:01 - BF1*null - BF2*015:01
17 BLB1*017:01 - BLB2*017:01 - BF1*004:04 - BF2*017:01
19 BLB1*012:01 - BLB2*012:02 - BF1*012:02 - BF2*015:02
21 BLB1*004:02 - BLB2*021:01 - BF1*004:02 - BF2*021:01
23 BLB1*005:02 - BLB2*023:01 - BF1*023:01 - BF2*002:02
24 BLB1*004:01 - BLB2*008:01 - BF1*004:03 - BF2*024:01
b.
BfBL BLB1 BLB2 BF1 BF2
6b BLB1*002:01 - BLB2*008:01 - BF1*006:01 - BF2*006:01
9 BLB1*0??:0? - BLB2*0??:0? - BF1*009:02 - BF2*009:01
9:02 BLB1*004:03 - BLB2*033:01 - BF1*004:04 - BF2*009:02
17:02 BLB1*0??:0? - BLB2*017:01 - BF1*030:01 - BF2*017:02
17:03 BLB1*0??:0? - BLB2*0??:0? - BF1*004:04 - BF2*017:03
24b BLB1*0??:0? - BLB2*036:01 - BF1*030:01 - BF2*024:01
30 BLB1*030:01 - BLB2*030:01 - BF1*006:02 - BF2*030:01
31 BLB1*031:01 - BLB2*031:01 - BF1*031:01 - BF2*031:01
32 BLB1*032:01 - BLB2*032:01 - BF1*004:02 - BF2*032:01
33 BLB1*009:01 - BLB2*034:01 - BF1*null - BF2*033:01
34 BLB1*004:01 - BLB2*0??:0? - BF1*023:02 - BF2*034:01
36 BLB1*0??:0? - BLB2*0??:0? - BF1*023:01 - BF2*036:01
36:02 BLB1*0??:0? - BLB2*0??:0? - BF1*006:01 - BF2*036:02
38 BLB1*109 - BLB2*109 - BF1*023:02 - BF2*038:01
39 BLB1*005:01-2 - BLB2*005:02 - BF1*004:01-3 - BF2*039:01
40 BLB1*033:01 - BLB2*035:01 - BF1*023:03 - BF2*040:01


a.
BLB1
BLB2
Conventional BLB
gene organisation
6 5 4 3 2 1
1 2 3 4 5 6
TAPASIN
BLB1
BLB2
6 5 4 3 2 1
1 2 3 4 5 6
TAPASIN
b.
BLB1
BLB2
Conventional BLB
gene organisation
6 5 4 3 2 1
1 2 3 4 5 6
TAPASIN
BLB1
BLB2
6 5 4 3 2 1
1 2 3 4 5 6
TAPASIN
or
BLB1
BLB2
6 5 4 3 2 1
1 2 3 4 5 6
TAPASIN

Online Resources 1, 2, 3, 4, 5, and 6 for

A potential nomenclature for the immunopolymorphism database (IPD) of chicken MHC genes: progress and problems

Hassnae Afrache[1], Clive A. Tregaskes[1] and Jim Kaufman[1,2,*]

[1]University of Cambridge, Department of Pathology, Tennis Court Road, Cambridge, CB2 1QP, U. K.

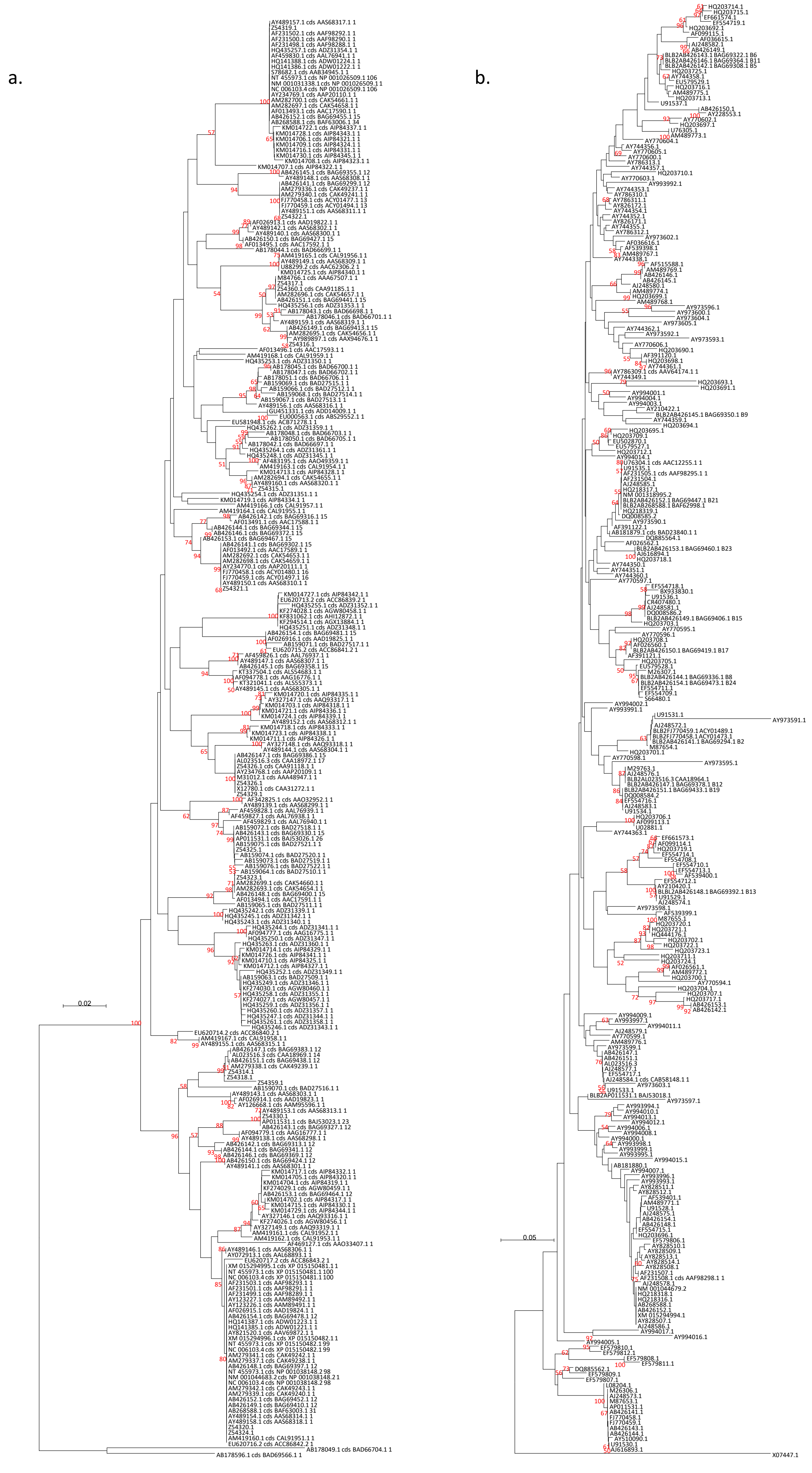
[2]University of Cambridge, Department of Veterinary Medicine, Madingley Road, Cambridge, CB2 0ES

*corresponding author, jfk31@cam.ac.uk

current email addresses

Hassnae Afrache, ha395@cam.ac.uk
Clive A. Tregaskes, ct383@cam.ac.uk
Jim Kaufman, jfk31@cam.ac.uk

Key words: BF-BL region, BF1, BF2, BLB1, BLB2, recombination


a.
b.
0.02
0.05

Online Resource 1. Phylogenetic trees for nucleotide sequences of exons encoding MHC peptide-binding domains from all chicken MHC-like sequences found in GenBank. a. exons 2 and 3 of chicken class I sequences; b. exon 2 of chicken class II B sequences. Sequences were found in two ways from the non-redundant GenBank database [National Center for Biotechnology Information (NCBI); www.ncbi.nlm.nih.gov/nuccore/] accessed in January 2017: by keywords (chicken AND MHC I AND gallus gallus, gallus gallus AND class I AND chromosome 16, g.gallus AND MHC class I, chicken AND MHC I AND gallus gallus AND chromosome 16), and by BLAST search using default parameters for nucleotide sequences (except using 1000 hits) and using AB426141 for BF and AB426141 for BLB as query sequences. Neighbour joining (NJ) trees were created by MEGA7 [Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molec Biol Evol 33:1870-1874] using Tamura-Nei method [Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molec Biol Evol 4:406-425]. Genetic distances are indicated with bars; red numbers are bootstrap values (percentages) for those nodes that reach significance from 500 replications; names at the tips are the GenBank accession number (followed by the accession number for the CDS protein sequence of some entries)

New name	Old names	Accession numbers			References
BF1*002:01:01	**BF1*0201**	**AM279336**	**whole**	**gene**	**Shaw et al 2007**
	BF1*0201	**AM279340**	**whole**	**gene**	**Shaw et al 2007**
	BF1*0201	**AB426141**	**whole**	**gene**	**Hosomichi et al 2008**
	BFCC9b-BFCC9-2	AY489151-AY489174	partial	gene	Lima-Rosa et al 2004
	B2m	**Z54322**	**partial**	**cDNA**	**Wallny et al 2006**
BF1*004:01:01	**BF1*0401**	**AM279337**	**whole**	**gene**	**Shaw et al 2007**
	BF1*0401	**AM279341**	**whole**	**gene**	**Shaw et al 2007**
	BF1*1301	**AB426148**	**whole**	**gene**	**Hosomichi et al 2008**
	BF1*JF2	AM419160	partial	gene	Worley et al 2008
	B4m	**Z54324**	**partial**	**cDNA**	**Wallny et al 2006**
BF1*004:02:01	**BF1*1501**	**AB426149**	**whole**	**gene**	**Hosomichi et al 2008**
	BF1*2101	**AM279339**	**whole**	**gene**	**Shaw et al 2007**
	BF1*2101	**AM279342**	**whole**	**gene**	**Shaw et al 2007**
	BF1*2101	**AB426152**	**whole**	**gene**	**Hosomichi et al 2008**
	BF1*21	**AY821520**	**whole**	**cDNA**	**Dalgaard et al 2005**
	BF1*W1	HQ141385	whole	cDNA	Kjaerup and Juul-Madsen, direct submission 2010
	BF1*131	HQ141387	whole	cDNA	Kjaerup and Juul-Madsen, direct submission 2010
	BF*C5b-BF*C5b	AY123227-AY128692	partial	gene	Livant et al 2004
	BFCC2b-BFCC15-2	AY489158-AY489181	partial	gene	Lima-Rosa et al 2004
	BFCC14b-BFCC12	AY489154-AY489177	partial	gene	Lima-Rosa et al 2004
	B-F200minor	AF231503	partial	cDNA	Juul-Madsen et al 2000
	B-F201minor	AF231501	partial	cDNA	Juul-Madsen et al 2000
	B-FW1minor	AF231499	partial	cDNA	Juul-Madsen et al 2000
	BA12	AF026915	partial	cDNA	Li et al 1999
	B21m	**Z54320**	**partial**	**cDNA**	**Wallny et al 2006**
BF1*004:03:01	**BF1*2401**	**AB426154**	**whole**	**gene**	**Hosomichi et al 2008**
BF1*004:04:01	**BF1*1701**	**AB426150**	**whole**	**gene**	**Hosomichi et al 2008**
	BFCC3b-BFCC3-2	AY489141-AY489164	partial	gene	Lima-Rosa et al 2004
BF1*004:05:01	BFCC6b-BFCC6-2	AY489146-AY489169	partial	gene	Lima-Rosa et al 2004
	BF*C7b	AY072913	partial	gene	Livant et al 2004
BF1*005:01:01	**BF1*0501**	**AB426142**	**whole**	**gene**	**Hosomichi et al 2008**
BF1*005:02:01	**BF1*0801**	**AB426144**	**whole**	**gene**	**Hosomichi et al 2008**
	BF1*1101	**AB426146**	**whole**	**gene**	**Hosomichi et al 2008**
BF1*006:01:01	**BF1*0601**	**AB426143**	**whole**	**gene**	**Hosomichi et al 2008**
	BF6M	**Z54330**	**partial**	**cDNA**	**Wallny et al 2006**
	BFCC10b-BFCC11	AY489153-AY489176	partial	gene	Lima-Rosa et al 2004
BF1*006:02:01	BF*C1b	AF094779	partial	gene	Livant et al 2004
	BFCC1b-BFCC1	AY489138-AY489161	partial	gene	Lima-Rosa et al 2004
BF1*009:01:01	**BF1*0901**	**AB426145**	**whole**	**gene**	**Hosomichi et al 2008**
BF1*009:02:01	BF*CC7b-BFCC7-2	AY489148-AY489171	partial	gene	Lima-Rosa et al 2004
BF1*012:01:01	**BF2*1202**	**AL023516**	**whole**	**gene**	**Kaufman et al 1999**
	BF1*1201	**AB426147**	**whole**	**gene**	**Shiina et al 2007**
	BF12m	**Z54314 with T4S**	**partial**	**cDNA**	**Wallny et al 2006**
BF1*012:02:01	**BF1*1902**	**AM279338**	**whole**	**gene**	**Shaw et al 2007**
	BF1*1901	**AB426151**	**whole**	**gene**	**Hosomichi et al 2008**
	BF19m	**Z54318 with T4S**	**partial**	**cDNA**	**Wallny et al 2006**
BF1*023:01:01	**BF1*2301**	**AB426153**	**whole**	**gene**	**Hosomichi et al 2008**
	BF*H8b	AY327146	partial	gene	Livant et al 2004
BF1*023:02:01	BF1*JF4	AM419161	partial	gene	Worley et al 2008
BF1*023:03:01	BF1*JF6	AM419162	partial	gene	Worley et al 2008
BF1*030:01:01	BA9-1	AF026914	partial	gene	Li et al 1999
	BF*A1b	AY126668	partial	gene	Livant et al 2004
	BFCC4b-BFCC4-2	AY489143-AY489166	partial	gene	Lima-Rosa et al 2004
BF1*031:01:01	BF*C2vb	AF469127	partial	gene	Livant et al 2004

Online Resource 2. Tables of potential BF and BLB names with accession numbers and citations for standard haplotypes and literature alleles; bold entries are for sequences from the standard haplotypes (page 1 of 8).

New name	Old names	Accession numbers			References
BF2*002:01:01	**BF2*0201**	**AM282692**	**whole**	**gene**	**Shaw et al 2007**
	BF2*0201	**AM282698**	**whole**	**gene**	**Shaw et al 2007**
	BF2*0201	**AB426141**	**whole**	**gene**	**Hosomichi et al 2008**
	BFIV2	AF013492	whole	cDNA	Hunt and Fulton 1998
	BFCC9a-BFCC9-1	AY489150-AY489173	partial	gene	Lima-Rosa et al 2004
	B2M	**Z54321**	**partial**	**cDNA**	**Wallny et al 2006**
BF2*002:02:01	**BF2*2301**	**AB426153**	**whole**	**gene**	**Hosomichi et al 2008**
BF2*004:01:01	**BF2*0401**	**AM282693**	**whole**	**gene**	**Shaw et al 2007**
	BF2*0401	**AM282699**	**whole**	**gene**	**Shaw et al 2007**
	BF2*1301	**AB426148**	**whole**	**gene**	**Hosomichi et al 2008**
	B4M	**Z54323**	**partial**	**cDNA**	**Wallny et al 2006**
BF2*005:01:01	**BF2*0501**	**AB426142**	**whole**	**gene**	**Hosomichi et al 2008**
	BFIV5	**AF013491**	**whole**	**cDNA**	**Hunt and Fulton 1998**
BF2*005:02:01	**BF2*0801**	**AB426144**	**whole**	**gene**	**Hosomichi et al 2008**
	BF2*1101	**AB426146**	**whole**	**gene**	**Hosomichi et al 2008**
BF2*006:01:01	**BF2*0601**	**AB426143**	**whole**	**gene**	**Hosomichi et al 2008**
	BF6m	**Z54325**	**partial**	**cDNA**	**Wallny et al 2006**
BF2*009:01:01	**BF2*0901**	**AB426145**	**whole**	**gene**	**Hosomichi et al 2008**
	BFCC7a-BFCC7-1	AY489147-AY489170	partial	gene	Lima-Rosa et al 2004
	Fayoumi	AF459826	partial	gene	Liu et al 2002
BF2*009:02:01	BFCC6a-BFCC6-1	AY489145-AY489168	partial	gene	Lima-Rosa et al 2004
	BC7	AF094778	partial	cDNA	Livant et al 2001
BF2*012:01:01	**B12**	**M31012**	**whole**	**gene**	**Kroemer et al 1990**
	BF1*1201	**AL023516**	**whole**	**gene**	**Kaufman et al 1999**
	BF2*1201	**AB426147**	**whole**	**gene**	**Hosomichi et al 2008**
	F10	**X12780**	**whole**	**cDNA**	**Guillemot 1988**
	BF12M	**Z54329**	**partial**	**cDNA**	**Wallny et al 2006**
BF2*014:01:01	**BF2*1401**	**AM282694**	**whole**	**gene**	**Shaw et al 2007**
	BFCC12a-BFCC17	AY489160-AY489183	partial	gene	Lima-Rosa et al 2004
BF2*015:01:01	**BF2*1501**	**AM282695**	**whole**	**gene**	**Shaw et al 2007**
	BF2*1501	**AB426149**	**whole**	**gene**	**Hosomichi et al 2008**
	BFIV15	**L28958**	**whole**	**cDNA**	**Hunt et al 1994**
	BF15M	**Z54316**	**partial**	**cDNA**	**Wallny et al 2006**
BF2*015:02:01	**BF2*1902**	**AM282696**	**whole**	**gene**	**Shaw et al 2007**
	BF2*1901	**AB426151**	**whole**	**gene**	**Hosomichi et al 2008**
	B19	**M84766**	**whole**	**cDNA**	**Kaufman et al 1992**
	BF19M	**Z54317**	**partial**	**cDNA**	**Wallny et al 2006**
BF2*015:03:01	BFCC11a-BFCC16	AY489159-AY489182	partial	gene	Lima-Rosa et al 2004
BF2*017:01:01	**BF2*1701**	**AB426150**	**whole**	**gene**	**Hosomichi et al 2008**
	BFIV17	**AF013495**	**whole**	**cDNA**	**Hunt and Fulton 1998**
BF2*017:02:01	BA1-1	AF026913	partial	cDNA	Li et al 1999
	BFCC4a-BFCC4-1	AY489142-AY489165	partial	gene	Lima-Rosa et al 2004
BF2*017:03:01	BFCC3a-BFCC3-1	AY489140-AY489163	partial	gene	Lima-Rosa et al 2004
BF2*018:01:01	**BFIV18**	**AF013496**	**whole**	**cDNA**	**Hunt and Fulton 1998**
BF2*021:01:01	**BF2*2101**	**AM282697**	**whole**	**gene**	**Shaw et al 2007**
	BF2*2101	**AM282700**	**whole**	**gene**	**Shaw et al 2007**
	BF2*2101	**AB426152**	**whole**	**gene**	**Hosomichi et al 2008**
	BFIV21	**S78682**	**whole**	**cDNA**	**Fulton et al 1995**
	BFIV21	**AF013493**	**whole**	**cDNA**	**Hunt and Fulton 1998**
	BF2*131	HQ141388	whole	cDNA	Kjaerup and Juul-Madsen, direct submission 2010
	BF2*W1	HQ141386	whole	cDNA	Kjaerup and Juul-Madsen, direct submission 2010
	BFCC2a-BFCC15-1	AY489157-AY489180	partial	gene	Lima-Rosa et al 2004
	Spanish	AF459830	partial	cDNA	Liu et al 2002
	BF21M	**Z54319**	**partial**	**cDNA**	**Wallny et al 2006**
BF2*024:01:01	**BF2*2401**	**AB426154**	**whole**	**gene**	**Hosomichi et al 2008**
	BA9-2	AF026916	partial	cDNA	Li et al 1999

New name	Old names	Accession numbers			References
BF2*030:01:01	BF*C1a	AF342825	partial	gene	Livant et al 2004
	BFCC1a-BFCC2a	AY489139-AY489162	partial	gene	Lima-Rosa et al 2004
BF2*031:01:01	BC2v	AF094777	partial	gene	Livant et al 2001
BF2*032:01	BF*A12a	AF483195	partial	gene	Livant et al 2004
	BF2*JF1	AM419163	partial	gene	Worley et al 2008
BF2*033:01:01	BFCC8a-BFCC8	AY489149-AY489172	partial	gene	Lima-Rosa et al 2004
	BF2*JF5	AM419165	partial	gene	Worley et al 2008
	BA4var	U88299	partial	cDNA	Li et al 1999
BF2*034:01:01	BFCC13a-BFCC13	AY489155-AY489178	partial	gene	Lima-Rosa et al 2004
	BF2*JF8	AM419167	partial	gene	Worley et al 2008
BF2*035:01:01	BF*J3a	AY327148	partial	gene	Livant et al 2004
	BFCC5a-BFCC5	AY489144-AY489167	partial	gene	Lima-Rosa et al 2004
BF2*036:01:01	BF*H8a	AY327147	partial	gene	Livant et al 2004
BF2*036:02:01	BFCC10a-BFCC10	AY489152-AY489175	partial	gene	Lima-Rosa et al 2004
BF2*037:01:01	BFCC14a-BFCC14	AY489156-AY489179	partial	gene	Lima-Rosa et al 2004
BF2*038:01:01	BF2*JF3	AM419164	partial	gene	Worley et al 2008
BF2*039:01:01	BF2*JF7	AM419166	partial	gene	Worley et al 2008
BF2*040:01:01	BF2*JF9	AM419168	partial	gene	Worley et al 2008

Online Resource 2, continued (page 3 of 8).

New name	Old name	Accession number			References
BLB1*002:01:01	**BLB1-B2**	**AB426141.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLB1-BR4	**FJ770458.1**	**whole**	**gene**	**Goto et al 2009**
	BLB1-BR2	**FJ770459.1**	**whole**	**gene**	**Goto et al 2009**
	B-LB2 minor	**AJ248573.1**	**partial**	**gene**	**Jacob et al 2000**
	B2b_B-LBI	**U91530.1**	**partial**	**cDNA**	**Pharr et al 1998**
	BLB1-B8	**AB426144.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLB1 WLA	AP011531.1	whole	gene	Suzuki et al 2012
BLB1*002:01:02	**BLB1-B6**	**AB426143.1**	**whole**	**gene**	**Hosomichi et al 2008**
BLB1*004:01:01	**B-LB4 minor**	**AJ248575.1**	**partial**	**gene**	**Jacob et al 2000**
	BLB1-B13	**AB426148.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLB1-B24	**AB426154.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLB1*JF10	AM489771.1	partial	gene	Worley et al 2008
BLB1*004:02:01	**BLB1-B21**	**AB426152.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B-LB21 minor	**AJ248586.1**	**partial**	**gene**	**Jacob et al 2000**
	BLBI	**AB268588.1**	**whole**	**gene**	**Shiina et al 2007**
	BLB1*131	HQ218316.1	partial	cDNA	Kjaerup and Juul-Madsen, direct submission 2010
	BLB1*W1	HQ218318.1	partial	cDNA	Kjaerup and Juul-Madsen, direct submission 2010
	GSP-BLB1	AB181880.1	partial	cDNA	Hosomichi, direct submission 2004
BLB1*004:03:01	B-L beta minor-BC7	AF539401.1	partial	gene	Livant and Ewald, direct submission 2002
BLB1*004:04:01	B-LW1minor	AF231507.1	partial	cDNA	Juul-Madsen et al 2000
	B200minor	AF231508.1	partial	cDNA	Juul-Madsen et al 2000
BLB1*005:01:01	**BLB1-B5**	**AB426142.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLB1*JF9	AM489770.1	partial	gene	Worley et al 2008
BLB1*005:02:01	**BLB1-B23**	**AB426153.1**	**whole**	**gene**	**Hosomichi et al 2008**
BLB1*005:03:01	B-L beta minor-BA9	AY702658.1	partial	gene	Livant and Ewald, direct submission 2004
BLB1*009:01:01	**BLB1-B9**	**AB426145.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLB1*JF7	AM489769.1	partial	gene	Worley et al 2008
	B-L beta minor-BA4v	AF515588.1	partial	gene	Livant and Ewald, direct submission 2002
BLB1*009:01:02	**BLB1-B11**	**AB426146.1**	**whole**	**gene**	**Hosomichi et al 2008**
BLB1*009:02:01	**B-LB14 minor**	**AJ248580.1**	**partial**	**gene**	**Jacob et al 2000**
BLB1*012:01:01	**BLb1**	**AL023516.3**	**whole**	**gene**	**Kaufman et al 1999**
	BLB1-B12	**AB426147.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B-LB12 minor	**AJ248577.1**	**partial**	**gene**	**Jacob et al 2000**
	BLB1-B19	**AB426151.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B-LB19 minor	**AJ248584.1**	**partial**	**gene**	**Jacob et al 2000**
	B19a_B-LBI	**U91533.1**	**partial**	**cDNA**	**Pharr et al 1998**
BLB1*015:01:01	**BLB1-B15**	**AB426149.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B-LB15 minor	**AJ248582.1**	**partial**	**gene**	**Jacob et al 2000**
BLB1*015:02:01	BA11-b allele	AF036615.1	partial	gene	Zheng et al 1999
BLB1*017:01:01	**BLB1-B17**	**AB426150.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B-L beta minor-BA1	AY228553.1	partial	gene	Livant and Ewald, direct submission 2003
BLB1*030:01:01	B-L beta	M87655.1	whole	cDNA	Pharr et al 1993
	B-L beta minor-BC1	AF539399.1	partial	gene	Livant and Ewald, direct submission 2002
BLB1*031:01:01	B-L beta minor-BC2v	AF539400.1	partial	gene	Livant and Ewald, direct submission 2002
BLB1*032:01:01	BLB1*JF2	AM489767.1	partial	gene	Worley et al 2008
	B-L beta minor-BA12	AF539398.1	partial	gene	Livant and Ewald, direct submission 2002
BLB1*032:02:01	BA12-b allele	AF036616.1	partial	gene	Zheng et al 1999
BLB1*033:01:01	BLB1*JF5	AM489768.1	partial	gene	Worley et al 2008
BLB*109	BLB*JF3	AM489776.1	partial	gene	Worley et al 2008

Online Resource 2, continued (page 4 of 8).

New name	Old name	Accession number			References
BLB2*002:01:01	**BLB-B2**	**AB426141.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLB2-BR4	**FJ770458.1**	**whole**	**gene**	**Goto et al 2009**
	BLB2-BR2	**FJ770459.1**	**whole**	**gene**	**Goto et al 2009**
	B-LB2 major	**AJ248572.1**	**partial**	**gene**	**Jacob et al 2000**
	B2a_B-LBII	**U91531.1**	**partial**	**cDNA**	**Pharr et al 1998**
BLB2*004:01:01	**B-LB4 major**	**AJ248574.1**	**partial**	**gene**	**Jacob et al 2000**
BLB2*004:02:01	**BLB2-B13**	**AB426148.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B13a_B-LBII	**U91529.1**	**partial**	**cDNA**	**Pharr et al 1998**
	B50 GB-1	AY210420.1	partial	gene	Zhou and Lamont 2003
BLB2*005:01:01	**BLB2-B5**	**AB426142.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLB2-B6	**AB426143.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B11	**AB426146.1**	**whole**	**gene**	**Hosomichi et al 2008**
BLB2*005:02:01	BLB2*JF8	AM489775.1	partial	gene	Worley et al 2008
BLB2*005:03:01	B-LB2*1001	AY744358.1	partial	gene	Xu et al 2007
BLB2*008:01:01	**BLB2-B8**	**AB426144.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLB2 WLA	AP011531.1	whole	gene	Suzuki et al 2012
BLB2*008:01:02	**BLB2-B24**	**AB426154.1**	**whole**	**gene**	**Hosomichi et al 2008**
BLB2*009:01:01	**BLB2-B9**	**AB426145.1**	**whole**	**gene**	**Hosomichi et al 2008**
BLB2*012:01:01	**BLb2**	**AL023516.3**	**whole**	**gene**	**Kaufman et al 1999**
	BLB2-B12	**AB426147.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B-LBII-beta	**M29763.1**	**whole**	**gene**	**Zoorob et al 1990**
	B-LB12 major	**AJ248576.1**	**partial**	**gene**	**Jacob et al 2000**
BLB2*012:02:01	**BLB-B19**	**AB426151.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B-LB-B19	**DQ008584.2**	**whole**	**cDNA**	**Niemiec et al 2006**
	B-LB19 major	**AJ248583.1**	**partial**	**gene**	**Jacob et al 2000**
	B19b_B-LBII	**U91534.1**	**partial**	**cDNA**	**Pharr et al 1998**
BLB2*014:01:01	**B-LB14 major**	**AJ248579.1**	**partial**	**gene**	**Jacob et al 2000**
BLB2*015:01:01	**B15**	**AB426149.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B-LB15 major	**AJ248581.1**	**partial**	**gene**	**Jacob et al 2000**
	B-LB-15I5	**DQ008586.2**	**whole**	**cDNA**	**Niemiec and Sharif, direct submission 2005**
BLB2*017:01:01	**BLB2-B17**	**AB426150.1**	**whole**	**gene**	**Hosomichi et al 2008**
	B-L beta allele A1	AF026560.1	partial	cDNA	Li et al 1999
BLB2*021:01:01	**B21**	**AB426152.1**	**whole**	**gene**	**Hosomichi et al 2008**
	BLBII	**AB268588.1**	**whole**	**gene**	**Shiina et al 2007**
	B-LB21 major	**AJ248585.1**	**partial**	**gene**	**Jacob et al 2000**
	B21_B-LBII	**U91535.1**	**partial**	**cDNA**	**Pharr et al 1998**
	BLB2*131	HQ218317.1	partial	cDNA	Kjaerup and Juul-Madsen, direct submission 2010
	BLB2*W1	HQ218319.1	partial	cDNA	Kjaerup and Juul-Madsen, direct submission 2010
	B-LW1major	AF231504.1	partial	cDNA	Juul-Madsen et al 2000
	B-L201major	AF231505.1	partial	cDNA	Juul-Madsen et al 2000
	B-L200major	AF231506.1	partial	cDNA	Juul-Madsen et al 2000
	BA4	U76304.1	partial	cDNA	Li et al 1997
	GSP-BLB2	AB181879.1	partial	cDNA	Hosomichi, direct submission 2004
BLB2*023:01:01	**BLB2-B23**	**AB426153.1**	**whole**	**gene**	**Hosomichi et al 2008**
BLB2*030:01:01	BC1	AF099113.1	partial	cDNA	Livant et al 2001
	B-L beta SPAFAS line 11	U02881.1	partial	cDNA	Pharr, direct submission 1993
	B-LB BQ355	HQ203706.1	whole	cDNA	Chen et al 2012
BLB2*031:01:01	BC2v	AF099114.1	partial	cDNA	Livant et al 2001
	B-LB BH478	HQ203719.1	whole	cDNA	Chen et al 2012
BLB2*032:01:01	BA12	AF026561.1	partial	cDNA	Li et al 1999
	BLB2*JF1	AM489772.1	partial	gene	Worley et al 2008
BLB2*033:01:01	BL-beta BC7	AF099115.1	partial	cDNA	Livant et al 2001
BLB2*034:01:01	BA4v	U76305.1	partial	cDNA	Li et al 1997
	BLB2*JF4	AM489773.1	partial	gene	Worley et al 2008
BLB2*035:01:01	BLB2*JF6	AM489774.1	partial	gene	Worley et al 2008
	B-LB BW462	HQ203699.1	whole	cDNA	Chen et al 2012
BLB2*036:01:01	BA9	AF026562.1	whole	cDNA	Li et al 1999
BLB*109	BLB*JF3	AM489776.1	partial	gene	Worley et al 2008

Citations:

Chen F, Pan L, Chao W, Dai Y, Yu W (2012) Character of chicken polymorphic major histocompatibility complex class II alleles of 3 Chinese local breeds. Poult Sci 91:1097-1104. doi: 10.3382/ps.2011-02007. PubMed PMID: 22499866.

Dalgaard TS, Vitved L, SkjÃ¸dt K, Thomsen B, Labouriau R, Jensen KH,

Juul-Madsen HR (2005) Molecular characterization of major histocompatibility complex class I (B-F) mRNA variants from chickens differing in resistance to Marek's disease. Scand J Immunol 62:259-270. PubMed PMID: 16179013

Fulton JE, Thacker EL, Bacon LD, Hunt HD (1995) Functional analysis of avian class I (BFIV) glycoproteins by epitope tagging and mutagenesis in vitro. Eur J Immunol 25:2069-2076. PubMed PMID: 7621880.

Goto RM, Wang Y, Taylor RL Jr, Wakenell PS, Hosomichi K, Shiina T, Blackmore

CS, Briles WE, Miller MM (2009) BG1 has a major role in MHC-linked resistance to malignant lymphoma in the chicken. Proc Natl Acad Sci U S A 106:16740-16745. doi: 10.1073/pnas.0906776106.PubMed PMID: 19805366; PubMedCentral PMCID: PMC2757851.

Guillemot F, Billault A, Pourquié O, Béhar G, Chaussé AM, Zoorob R, Kreibich

G, Auffray C. (1998) A molecular map of the chicken major histocompatibility complex: the class II beta genes are closely linked to the class I genes and the nucleolar organizer. EMBO J 7:2775-2785. PubMed PMID: 3141149; PubMed CentralPMCID: PMC457068.

Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, Nishibori M,Inoko H, Hanzawa K, Shiina T (2008) Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. J Immunol 181:3393-3399. PubMed PMID: 18714011; PubMed Central PMCID: PMC2657362.

Hunt HD, Fulton JE (1998) Analysis of polymorphisms in the major expressed class I locus (B-FIV) of the chicken. Immunogenetics 47:456-467. PubMed PMID: 9553152.

Hunt HD, Pharr GT, Bacon LD (1994) Molecular analysis reveals MHC class I intra-locus recombination in the chicken. Immunogenetics 40:370-375. PubMed PMID: 7927541.

Jacob JP, Milne S, Beck S, Kaufman J (2000) The major and a minor class II beta-chain (B-LB) gene flank the Tapasin gene in the B-F /B-L region of the chicken major histocompatibility complex. Immunogenetics 51:138-147. PubMed PMID: 10663576.

Juul-Madsen HR, Dalgaard TS, Afanassieff M (2000) Molecular characterization of major and minor MHC class I and II genes in B21-like haplotypes in chickens. Anim Genet 31:252-261. PubMed PMID: 11086534.

Kaufman J, Andersen R, Avila D, Engberg J, Lambris J, Salomonsen J, Welinder K, Skjodt K (1992) Different features of the MHC class I heterodimer have evolved at different rates. Chicken B-F and beta 2-microglobulin sequences reveal invariant surface residues. J Immunol 148:1532-1546. PubMed PMID: 1538136.

Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923-925. PubMed PMID: 10553909.

Kroemer G, Zoorob R, Auffray C (1990) Structure and expression of a chicken MHC class I gene. Immunogenetics 31:405-409. PubMed PMID: 2370087.

Li L, Johnson LW, Ewald SJ (1997) Molecular characterization of major histocompatibility complex (B) haplotypes in broiler chickens. Anim Genet 28:258-267. PubMed PMID: 9345722.

Li L, Johnson LW, Livant EJ, Ewald SJ (1999) The MHC of a broiler chicken line: serology, B-G genotypes, and B-F/B-LB sequences. Immunogenetics 49:215-224. PubMed PMID: 9914335.

Lima-Rosa CA, Canal CW, Streck AF, Freitas LB, Delgado-Canedo A, Bonatto SL, Salzano FM (2004) B-F DNA sequence variability in Brazilian (blue-egg Caipira) chickens. Anim Genet 35:278-284. PubMed PMID: 15265066.

Liu W, Miller MM, Lamont SJ (2002) Association of MHC class I and class II gene polymorphisms with vaccine or challenge response to Salmonella enteritidis in young chicks. Immunogenetics 54:582-590. PubMed PMID: 12439621.

Livant EJ, Zheng D, Johnson LW, Shi W, Ewald SJ (2001) Three new MHC haplotypes in broiler breeder chickens. Anim Genet 32:123-131. PubMed PMID: 11493260.

Livant EJ, Brigati JR, Ewald SJ (2004) Diversity and locus specificity of chicken MHC B class I sequences. Anim Genet 35:18-27. PubMed PMID: 14731225.

Niemiec PK, Read LR, Sharif S. (2006) Synthesis of chicken major histocompatibility complex class II oligomers using a baculovirus expression system. Protein Expr Purif 46:390-400. PubMed PMID: 16236525.

Pharr GT, Bacon LD, Dodgson JB (1993) Analysis of B-L beta-chain gene expression in two chicken cDNA libraries. Immunogenetics 37:381-385. PubMed PMID:8428771.

Pharr GT, Dodgson JB, Hunt HD, Bacon LD (1998) Class II MHC cDNAs in 15I5 B-congenic chickens. Immunogenetics 47:350-354. PubMed PMID: 9510552.

Shaw I, Powell TJ, Marston DA, Baker K, van Hateren A, Riegert P, Wiles MV, Milne S, Beck S, Kaufman J (2007) Different evolutionary histories of the two classical class I genes BF1 and BF2 illustrate drift and selection within the stable MHC haplotypes of chickens. J Immunol 178:5744-5752. PubMed PMID: 17442958.

Shiina T, Briles WE, Goto RM, Hosomichi K, Yanagiya K, Shimizu S, Inoko H, Miller MM (2007) Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. J Immunol 178:7162-7172. PubMed PMID: 17513765.

Suzuki K, Kobayashi E, Yamashita H, Uenishi H, Churkina I, Plastow G, Hamasima N, Mitsuhashi T (2012) Structural analysis of MHC alleles in an RSV tumour regression chicken using a BAC library. Anim Genet 43:348-351. doi: 10.1111/j.1365-2052.2011.02247.x. PubMed PMID: 22486511.

Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, Skjodt K, Vainio O, Vilbois F, Wiles MV, Kaufman J (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci U S A 103:1434-1439. PubMed PMID: 16432226; PubMed Central PMCID: PMC1360531.

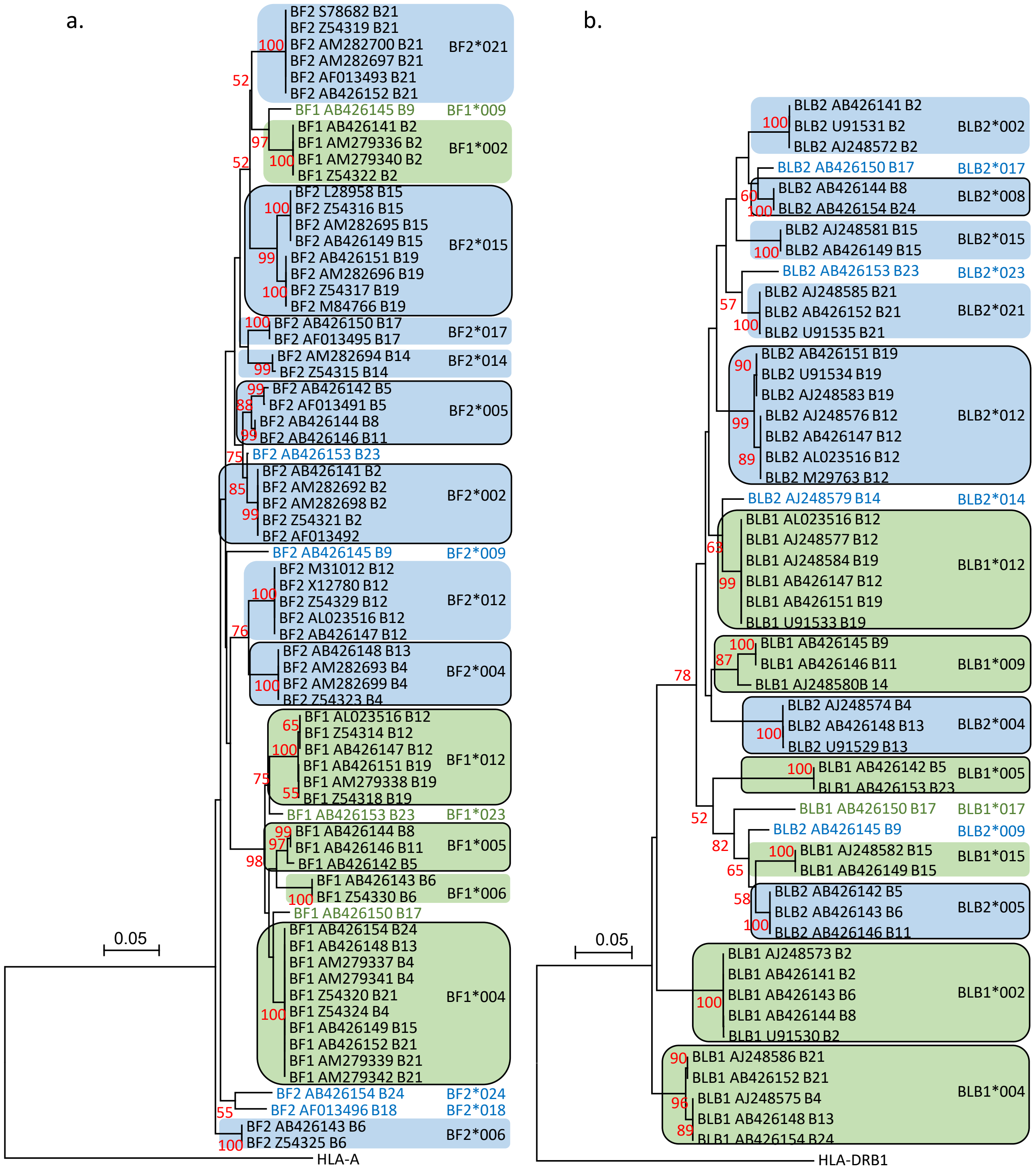
Worley K, Gillingham M, Jensen P, Kennedy LJ, Pizzari T, Kaufman J, Richardson DS (2008) Single locus typing of MHC class I and class II B loci in a population of red jungle fowl. Immunogenetics 60:233-247. doi: 10.1007/s00251-008-0288-0. PubMed PMID: 18389232.

Xu R, Li K, Chen G, Xu H, Qiang B, Li C, Liu B (2007) Characterization of genetic polymorphism of novel MHC B-LB II alleles in Chinese indigenous chickens. J Genet Genomics 34:109-118. PubMed PMID: 17469783.

Zheng D, O'Keefe G, Li L, Johnson LW, Ewald SJ (1999) A PCR method for typing B-L beta II family (class II MHC) alleles in broiler chickens. Anim Genet 30:109-119. PubMed PMID: 10376301.

Zhou H, Lamont SJ (2003) Chicken MHC class I and II gene effects on antibody response kinetics in adult chickens. Immunogenetics 55:133-140. doi: 10.1007/s00251-003-0566-9. PubMed PMID: 12743657.

Zoorob R, Béhar G, Kroemer G, Auffray C (1990) Organization of a functional chicken class II B gene. Immunogenetics 31:179-187. PubMed PMID: 1969383.



Online Resource 3. Phylogenetic trees for nucleotide sequences of exons encoding MHC peptide-binding domains from standard haplotypes. a. exons 2 and 3 of BF sequences (with the first 20 nucleotides of exon 2 and the last 23 nucleotides of exon 3 removed, corresponding to primers and other reasons for different lengths of sequence); b. exon 2 of BLB sequences (with the first 6-8 nucleotides removed, corresponding to primers and other reasons for different lengths of sequence). Genetic distances are indicated with bars; red numbers are bootstrap values (percentages) for those nodes that reach significance from 500 replications; names at the tips are the gene name, followed by the GenBank accession number, followed by the haplotype. Allele groups for BF1 and BLB1 (or BF2 and BLB2) are named, either in green (or blue) for single sequences or in black surrounded by green (or blue) background for clades with more than one sequence; the coloured background for clades with sequences from more than one haplotype are surrounded by a black line. Human sequences were used as outgroups; sequences for standard haplotypes were taken from the GenBank accession numbers in Online Resource 2; all other details are as in the legend to Online Resource 1.

a.

BF1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B15	B17	B19	B21	B23	B24
B2	ID														
B4	53	ID													
B5	50	18	ID												
B6	46	26	24	ID											
B8	46	16	4	20	ID										
B9	20	54	48	52	52	ID									
B11	46	16	4	20	0	52	ID								
B12	57	25	28	36	32	51	32	ID							
B13	53	0	18	26	16	54	16	25	ID						
B15	53	0	18	26	16	54	16	25	0	ID					
B17	53	12	17	25	15	52	15	29	12	12	ID				
B19	57	24	27	35	31	51	31	1	24	24	28	ID			
B21	53	0	18	26	16	54	16	25	0	0	12	24	ID		
B23	57	16	24	34	26	47	26	21	16	16	22	20	16	ID	
B24	53	0	18	26	16	54	16	25	0	0	12	24	0	16	ID

b.

BF2	B2	B4	B5	B6	B8	B9	B11	B12	B13	B14	B15	B17	B18	B19	B21	B23	B24
B2	ID																
B4	39	ID															
B5	18	40	ID														
B6	39	41	40	ID													
B8	12	39	9	37	ID												
B9	34	46	35	35	36	ID											
B11	12	39	9	37	0	36	ID										
B12	19	24	29	38	27	36	27	ID									
B13	39	0	40	41	39	46	39	24	ID								
B14	28	41	28	37	26	36	26	33	41	ID							
B15	36	53	38	39	37	35	37	52	53	25	ID						
B17	23	43	21	39	24	35	24	33	43	20	33	ID					
B18	36	35	34	32	37	38	37	35	35	25	41	31	ID				
B19	33	48	33	36	34	35	34	47	48	28	10	33	42	ID			
B21	30	43	30	41	28	40	28	45	43	34	35	31	44	27	ID		
B23	5	38	14	34	7	34	7	22	38	23	34	18	32	31	29	ID	
B24	35	44	34	34	39	40	39	34	44	41	52	34	29	51	49	34	ID

c.

BF2/BF1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B15	B17	B19	B21	B23	B24
B2	38	40	34	46	38	39	38	44	40	40	38	44	40	40	40
B4	47	40	53	51	49	48	49	49	40	40	46	49	40	44	40
B5	38	47	40	45	44	34	44	43	47	47	43	43	47	43	47
B6	35	41	43	35	39	37	39	50	41	41	42	50	41	43	41
B8	37	38	31	45	35	31	35	38	38	38	34	38	38	34	38
B9	42	42	41	49	43	40	43	46	42	42	49	46	42	45	42
B11	37	38	31	45	35	31	35	38	38	38	34	38	38	34	38
B12	48	43	43	47	47	48	47	45	43	43	49	45	43	41	43
B13	47	40	53	51	49	48	49	49	40	40	46	49	40	44	40
B14	46	49	42	42	46	41	46	47	49	49	40	47	49	51	49
B15	35	54	48	50	50	36	50	58	54	54	45	58	54	58	54
B17	45	46	40	38	44	39	44	50	46	46	40	50	46	45	46
B18	47	49	52	45	50	48	50	50	49	49	46	50	49	49	49
B19	35	55	47	49	51	34	51	58	55	55	48	58	55	56	55
B21	35	53	48	47	52	28	52	48	53	53	51	48	53	51	53
B23	36	35	29	41	33	34	33	39	35	35	33	39	35	35	35
B24	44	49	51	44	47	50	47	56	49	49	55	55	49	51	49

d.

BLB1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B14	B15	B17	B19	B21	B23	B24
B2	ID															
B4	26	ID														
B5	45	46	ID													
B6	0	26	45	ID												
B8	0	26	45	0	ID											
B9	37	31	35	37	37	ID										
B11	37	31	35	37	37	0	ID									
B12	33	25	35	33	33	19	19	ID								
B13	26	0	46	26	26	31	31	25	ID							
B14	30	31	31	30	30	7	7	18	31	ID						
B15	40	45	38	40	40	28	28	28	45	26	ID					
B17	38	46	33	38	38	34	34	32	46	31	23	ID				
B19	33	25	35	33	33	19	19	0	25	18	28	32	ID			
B21	24	2	46	24	24	33	33	25	2	33	45	44	25	ID		
B23	45	46	0	45	45	35	35	35	46	31	38	33	35	46	ID	
B24	26	0	46	26	26	31	31	25	0	31	45	46	25	2	46	ID

e.

BLB2	B2	B4	B5	B6	B8	B9	B11	B12	B13	B14	B15	B17	B19	B21	B23	B24
B2	ID															
B4	31	ID														
B5	34	36	ID													
B6	34	36	0	ID												
B8	12	31	32	32	ID											
B9	36	36	10	10	33	ID										
B11	34	36	0	0	32	10	ID									
B12	16	29	30	30	21	28	30	ID								
B13	31	0	36	36	31	36	36	29	ID							
B14	26	21	24	24	25	20	24	19	21	ID						
B15	21	35	34	34	14	35	34	26	35	25	ID					
B17	17	29	29	29	7	29	29	22	29	23	20	ID				
B19	16	31	30	30	19	28	30	2	31	21	24	20	ID			
B21	21	24	28	28	13	25	28	20	24	20	19	12	20	ID		
B23	29	28	33	33	24	30	33	24	28	25	24	19	24	12	ID	
B24	12	31	32	32	0	33	32	21	31	25	14	7	19	13	24	ID

f.

BLB2/BLB1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B14	B15	B17	B19	B21	B23	B24
B2	41	38	44	41	41	33	33	24	38	34	43	38	24	36	44	38
B4	39	42	40	39	39	21	21	25	42	28	38	40	25	40	40	42
B5	39	39	37	39	39	30	30	25	39	25	12	18	25	37	37	39
B6	39	39	37	39	39	30	30	25	39	25	12	18	25	37	37	39
B8	43	34	41	43	43	28	28	20	34	29	41	38	20	32	41	34
B9	35	37	41	35	35	32	32	25	37	27	14	24	25	35	41	37
B11	39	39	37	39	39	30	30	25	39	25	12	18	25	37	37	39
B12	35	30	42	35	35	26	26	18	30	26	38	34	18	28	42	30
B13	39	42	40	39	39	21	21	25	42	28	38	40	25	40	40	42
B14	30	27	36	30	30	18	18	9	27	17	29	29	9	27	36	27
B15	35	32	35	35	35	23	23	22	32	24	37	37	22	32	35	32
B17	38	33	36	38	38	27	27	21	33	23	39	36	21	31	36	33
B19	37	28	40	37	37	26	26	16	28	26	38	35	16	26	40	28
B21	34	34	34	34	34	21	21	20	34	18	33	34	20	32	34	34
B23	35	33	32	35	35	20	20	23	33	16	33	36	23	35	32	33
B24	43	34	41	43	43	28	28	20	34	29	41	38	20	32	41	34

Online resource 4. Distance matrices for nucleotide sequences of exons encoding MHC peptide-binding domains from standard haplotypes. Exons 2 and 3 of a. BF1 versus BF1 alleles, b. BF2 versus BF2 alleles, c. BF1 versus BF2 alleles; d. BLB1 versus BLB1 alleles, e. BLB2 versus BLB2 alleles, f. BLB1 versus BLB2 alleles. Sequences for standard haplotypes were taken from the GenBank accession numbers in Online Resource 2. Alignments were performed using MAFFT on-line [Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066; https://mafft.cbrc.jp/alignment/server/] and the results were pasted into Bioedit [Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98; https://softfamous.com/bioedit/] on a desktop computer; the command “Sequence difference count Matrix” under “Alignment” was used to generate the distance matrix, which was pasted into Microsoft Excel and then Powerpoint for producing the final figure. Highlights indicate amino acid differences for BF (or BLB) from Fig. 3 (for comparison to nucleotide differences in this figure): green, none; blue, 1 to 4 (1 or 2); yellow, 5 to 8 (3 or 4); ID, comparison between the same sequence.

a.

BF1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B15	B17	B19	B21	B23	B24
B2	ID														
B4	47	ID													
B5	43	17	ID												
B6	40	18	16	ID											
B8	39	15	4	12	ID										
B9	23	54	46	51	50	ID									
B11	39	15	4	12	0	50	ID								
B12	41	30	28	33	32	44	32	ID							
B13	47	0	17	18	15	54	15	30	ID						
B15	45	2	17	18	15	54	15	30	2	ID					
B17	47	11	19	20	17	53	17	33	11	9	ID				
B19	41	29	27	32	31	44	31	1	29	29	32	ID			
B21	45	2	17	18	15	54	15	30	2	0	9	29	ID		
B23	44	22	27	31	28	43	28	20	22	22	26	19	22	ID	
B24	47	2	19	20	17	56	17	32	2	2	11	31	2	24	ID

b.

BF2	B2	B4	B5	B6	B8	B9	B11	B12	B13	B14	B15	B17	B18	B19	B21	B23	B24
B2	ID																
B4	27	ID															
B5	17	32	ID														
B6	33	31	33	ID													
B8	15	31	5	33	ID												
B9	31	36	30	37	31	ID											
B11	15	31	5	33	0	31	ID										
B12	11	18	24	31	23	32	23	ID									
B13	27	0	32	31	31	36	31	18	ID								
B14	24	31	26	34	25	32	25	27	31	ID							
B15	27	35	31	33	30	33	30	35	35	21	ID						
B17	21	34	21	35	23	29	23	28	34	23	28	ID					
B18	44	44	47	47	48	50	48	44	44	43	48	45	ID				
B19	24	31	26	28	27	32	27	30	31	22	10	26	49	ID			
B21	24	31	23	35	24	32	24	32	31	28	28	26	52	22	ID		
B23	5	26	13	29	10	31	10	15	26	20	25	16	40	22	24	ID	
B24	29	33	32	29	35	38	35	30	33	36	39	31	46	38	37	29	ID

c.

BF2/BF1	B2	B4	B5	B6	B8	B9	B11	B12	B13	B15	B17	B19	B21	B23	B24
B2	36	41	37	43	41	42	41	31	41	39	40	31	39	34	41
B4	33	39	47	42	43	45	43	34	39	37	42	34	37	35	39
B5	36	43	39	42	43	41	43	30	43	41	42	30	41	36	43
B6	34	43	44	36	40	42	40	37	43	41	42	37	41	38	43
B8	35	38	34	43	38	39	38	29	38	36	37	29	36	31	38
B9	39	43	40	45	42	42	42	33	43	41	46	33	41	37	43
B11	35	38	34	43	38	39	38	29	38	36	37	29	36	31	38
B12	40	44	42	45	46	46	46	32	44	42	46	32	42	35	44
B13	33	39	47	42	43	45	43	34	39	37	42	34	37	35	39
B14	40	44	39	42	43	44	43	28	44	42	41	28	42	38	44
B15	34	43	41	42	43	39	43	33	43	43	42	33	43	38	45
B17	41	43	39	41	43	42	43	32	43	41	38	32	41	33	43
B18	57	60	62	57	60	62	60	51	60	60	62	51	60	56	62
B19	33	43	40	41	44	36	44	32	43	43	41	32	43	38	45
B21	31	46	40	40	44	35	44	33	46	44	44	33	44	39	46
B23	34	36	32	39	36	39	36	26	36	34	35	26	34	29	36
B24	41	44	46	39	42	53	42	38	44	42	46	38	42	39	44

d.

BLB1	B2	B5	B6	B8	B9	B11	B12	B13	B15	B17	B19	B21	B23	B24
B2	ID													
B5	23	ID												
B6	0	23	ID											
B8	0	23	0	ID										
B9	22	22	22	22	ID									
B11	22	22	22	22	0	ID								
B12	18	20	18	18	16	16	ID							
B13	17	25	17	17	20	20	17	ID						
B15	24	22	24	24	20	20	20	26	ID					
B17	23	19	23	23	20	20	21	27	16	ID				
B19	18	20	18	18	16	16	0	17	20	21	ID			
B21	16	24	16	16	21	21	16	1	25	26	16	ID		
B23	35	12	35	35	34	34	32	37	34	31	32	36	ID	
B24	17	25	17	17	20	20	17	0	26	27	17	1	37	ID

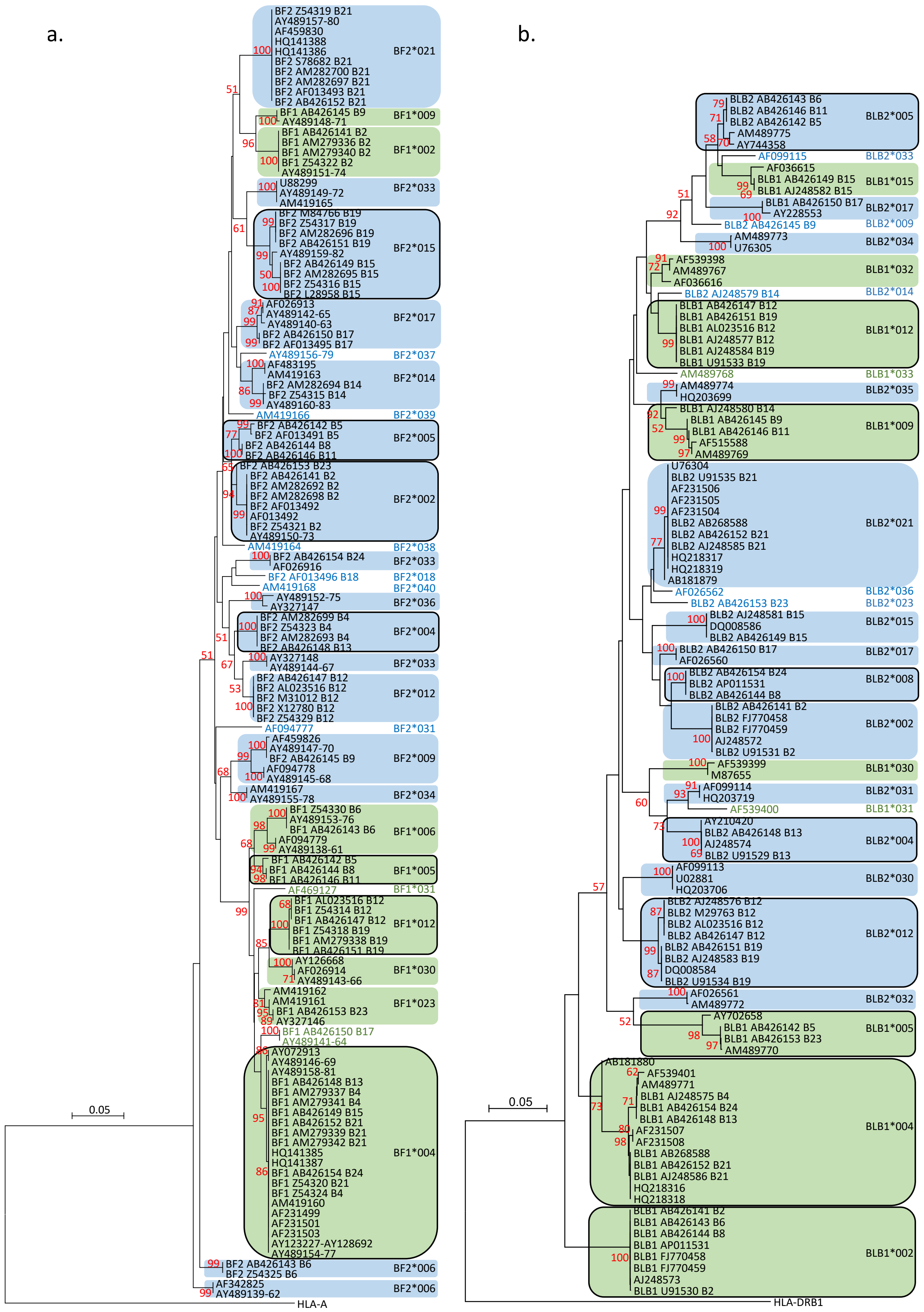
e.

BLB2	B2	B5	B6	B8	B9	B11	B12	B13	B15	B17	B19	B21	B23	B24
B2	ID													
B5	23	ID												
B6	23	0	ID											
B8	8	22	22	ID										
B9	22	11	11	20	ID									
B11	23	0	0	22	11	ID								
B12	11	20	20	14	17	20	ID							
B13	21	20	20	20	22	20	20	ID						
B15	15	22	22	11	23	22	18	20	ID					
B17	13	22	22	7	20	22	17	21	17	ID				
B19	11	20	20	12	17	20	2	20	16	15	ID			
B21	15	20	20	10	18	20	14	18	14	13	14	ID		
B23	18	21	21	15	19	21	15	18	17	14	15	9	ID	
B24	8	22	22	0	20	22	14	20	11	7	12	10	15	ID

f.

BLB2/BLB1	B2	B5	B6	B8	B9	B11	B12	B13	B15	B17	B19	B21	B23	B24
B2	24	25	24	24	22	22	15	21	25	24	15	20	37	21
B5	26	21	26	26	18	18	18	24	9	12	18	23	33	24
B6	26	21	26	26	18	18	18	24	9	12	18	23	33	24
B8	24	23	24	24	20	20	13	20	25	25	13	19	35	20
B9	22	23	22	22	20	20	17	21	12	19	17	20	35	21
B11	26	21	26	26	18	18	18	24	9	12	18	23	33	24
B12	22	23	22	22	17	17	13	18	24	22	13	17	35	18
B13	22	23	22	22	14	14	14	24	24	21	14	23	35	24
B15	19	22	19	19	18	18	15	21	25	24	15	20	34	21
B17	24	23	24	24	22	22	16	21	26	26	16	20	34	21
B19	23	21	23	23	17	17	11	16	24	22	11	15	33	16
B21	21	20	21	21	17	17	15	22	22	24	15	21	32	22
B23	20	20	20	20	14	14	16	18	22	24	16	19	32	18
B24	24	23	24	24	20	20	13	20	25	25	13	19	35	20

Online Resource 5. Distance matrices for amino acids of whole coding sequences (CDS) from standard haplotypes. a. BF1 versus BF1 alleles, b. BF2 versus BF2 alleles, c. BF1 versus BF2 alleles; d. BLB1 versus BLB1 alleles, e. BLB2 versus BLB2 alleles, f. BLB1 versus BLB2 alleles. All details as in legend to Online Resource 4, except that the names of sequences with indels are highlighted: grey, insertion (one amino acid in BF1, two amino acids in BF2); pink, deletion (five amino acids in BF1, 11 amino acids in BF2); orange, a combination of a one nucleotide deletion and a five nucleotide truncation leading to a frameshift at amino acid 247 in the transmembrane region (BLB1 from the B23 haplotype).



Online Resource 6. Phylogenetic trees for nucleotide sequences of exons encoding MHC peptide-binding domains from standard haplotypes and from sequences in the scientific literature. a. exons 2 and 3 of BF sequences (with the first 20 nucleotides of exon 2 and the last 23 nucleotides of exon 3 removed, corresponding to primers and other reasons for different lengths of sequence); b. exon 2 of BLB sequences (with the first 6-8 nucleotides removed, corresponding to primers and other reasons for different lengths of sequence). Sequences for standard haplotypes and from the scientific literature were taken from the GenBank accession numbers in Online Resource 2; all other details are as in the legends to Online Resources 1 and 3.